TITLE OF THE INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit under 35 USC §119 of provisional application USSN 60/239,924 filed on October 13, 2000 and of provisional application USSN 60/283,296 filed April 12, 2001, and claims benefit under 35 USC § 120 of USSN 90/910,813 which are hereby incorporated by reference in their entirety for all purposes.

FIELD OF INVENTION:

[0002] The present invention relates to the field of antibiotics, and more specifically to genes involved in the biosynthesis of ramoplanin. The invention provides recombinant methods and materials for producing ramoplanins by recombinant DNA technology.

BACKGROUND:

[0003] Ramoplanin is a naturally-occurring glycosylated lipodepsipeptide antibiotic active against Gram-positive aerobic and anaerobic bacteria. Ramoplanin kills Gram-positive bacteria by inhibiting one of the enzymes needed to construct the bacterial cell wall. Ramoplanin was first described as antibiotic A/16686 produced by fermentation of Actinoplanes sp. ATCC 33076, as described in U.S. Patent No. 4,303,646. It was subsequently found that three closely related components could be isolated from antibiotic A/16686, which components were named antibiotic A/16686 factors A1, A2, and A3 (Ciabatti et al., 1989, J. Antibiot (Tokyo), Vol. 42, No. 2, pp. 254-267). These substances as well as their preparation and uses are described in U.S. Patent No. 4,427,656. Three additional factors designated A'1, A'2, and A'3 were later shown to be present in the fermentation medium and were shown to differ from the respective parent components of the original complex by lacking one mannose unit from the glycosidic group (Gastaldo et al., 1992, J. Ind. Microbiol. Vol. 11, No. 1, pp. 13-18).

[0004] Ramoplanin consists of a mixture of three related polypeptides having a common cyclic depsipeptide core structure on which is carried a dimannosyl glycosidic group. The three forms of ramoplanin are differentiated by the presence of various acylamide moieties derived from 8-, 9-, or 10-carbon fatty acids that decorate the glycosylated depsipeptide core structure.

[0005] Depsipeptides are cyclic or branched peptides containing an ester linkage between a carboxylate group of the peptide and a terminal or side-chain hydroxyl group of the peptide. The ramoplanin depsipeptide core structure contains 17 amino acids.

The order of amino acids, from N-terminal to C-terminal, is as follows: amino acid 1: asparagine (Asn); amino acid 2: beta-hydroxyasparagine (HAsn); amino acid 3: 4-hydroxyphenylglycine (HPG); amino acid 4: ornithine (Orn); amino acid 5: threonine (Thr); amino acid 6: HPG; amino acid 7: HPG; amino acid 8: Thr; amino acid 9: phenylalanine (Phe); amino acid 10: Orn; amino acid 11: HPG; amino acid 12: Thr; amino acid 13: HPG; amino acid 14: glycine (Gly); amino acid 15: leucine (Leu); amino acid 16: alanine (Ala); amino acid 17: 3-chloro-4-hydroxyphenylglycine (CHPG). The peptide is cyclized by ester bond formation between the carboxylate group of the C-terminal CHPG and the hydroxyl group of HAsn. The N-terminus of Asn in position 1 is acylated by three different fatty acids, resulting in the three different components A1-A3. Two D-mannose sugars are attached to the HPG in position 11 by a hemiacetal bond.

[0006] Many low molecular weight peptides produced by bacteria are synthesized nonribosomally on large multifunctional proteins termed peptide synthetases. (Konz & Marahiel, 1999, Chem. Biol., Vol. 6, pp. R39-R48). Peptide synthetases contain repeated units that each recognize specific amino acids and catalyze their stepwise joining into a peptide chain. The identity of the amino acid recognized by a particular unit can be determined by comparison with other units of known specificity. In many peptide synthetases, there is a strict correlation between the order of repeated units in a peptide synthetase and the order in which the respective amino acids appear in the peptide product, making it possible to correlate peptides of known structure with putative genes encoding their synthesis, as demonstrated by the identification of the mycobactin biosynthetic gene cluster from the genome of Mycobacterium tuberculosis (Quadri et al., 1998, Chem. Biol. Vol. 5, pp. 631-645).

[0007] The repeating units of a peptide synthetase are composed of smaller units or "domains" that each carry out a specific role in the recognition, activation, modification and joining of amino acid precursors to form the peptide product. One type of domain, the adenylation (A) domain, is responsible for selectively recognizing and activating the amino acid that is to be incorporated by a particular unit of the peptide synthetase. The activated amino acid is joined to the peptide synthetase through another type of domain, the thiolation (T) domain, that is generally located adjacent to the A domain. Amino acids joined to successive units of the peptide synthetase are subsequently linked together by the formation of amide bonds catalyzed by another type of domain, the condensation (C) domain.

[0008] Although the structure of ramoplanin has been identified, there remains the need to obtain novel structures with new activities or enhanced properties. There is also a need to improve production of ramoplanin. Accordingly, there is a need for genetic information regarding the biosynthesis of ramoplanin.

SUMMARY OF THE INVENTION:

[0009] The present invention provides purified and isolated polynucleotide molecules that encode polypeptides of the ramoplanin biosynthetic pathway in microorganisms. In one form of the invention, polynucleotide molecules are selected from the contiguous DNA sequence (SEQ ID NO: 1) representing the full-length locus of the ramoplanin biosynthetic pathway and containing the 34 ORFs encoding the proteins forming the ramoplanin gene cluster. The amino acid sequence of the proteins is provided in SEQ ID NOS: 2 to 34. Structural and functional characterization is provided for the 32 ORFs.

[0010] Thus, in one aspect, the invention provides an isolated nucleic acid comprising a nucleic acid sequence selected from the group consisting of (a) nucleic acid encoding any of ramoplanin ORFs 1 to 33 (SEQ ID NOS: 2 to 34); (b) a nucleic acid encoding a polypeptide encoded by any of ramoplanin ORFs 1 to 33 (SEQ ID NOS: 2 to 34); and (c) a nucleic acid encoding a polypeptide that is at least 75%, preferably 80%, more preferably 85%, still more preferably 90% and most preferably 95% or more identical in amino acid sequence to a polypeptide of ramoplanin ORFs 4, 5, 9 to 19, 22 to 26, 29, 30 and 31 (SEQ ID NOS: 5, 6, 10 to 20, 23 to 27, 30, 31 and 32).

[0011] Certain embodiments of the invention specifically exclude one or more of ORFs 1 to 32, most notably ORFs 1, 2, 3, 6, 7, 8, 20, 21, 27, 28, 31 and 32 (SEQ ID NOS: 2, 3, 4, 7, 8, 9, 221, 22, 28, 29, 32 and 33) although other ORFs can be excluded without departing from the scope of the invention. Thus, another embodiment of the invention provides an isolated nucleic acid comprising a nucleic acid sequence selected from the group consisting of: (a) a nucleic acid encoding any of ramoplanin ORFs 4, 5, 9 to 19, 22 to 26, 29, 30 and 31 (SEQ ID NOS: 5, 6, 10 to 20, 23 to 27, 30, 31 and 32); (b) a nucleic acid encoding a polypeptide encoded by any of ramoplanin ORFs 4, 5, 9 to 19, 22 to 26, 29, 30 and 31 (SEQ ID NOS: 5, 6, 10 to 20, 23 to 27, 30, 31 and 32); and (c) a nucleic acid encoding a polypeptide that is at least 75%, preferably 80%, more preferably 85%, still more preferably 90% and most preferably 95% or more identical in amino acid sequence to a polypeptide of ramoplanin ORFs 4, 5, 9 to 19, 22 to 26, 29, 30 and 31 (SEQ ID NOS: 5, 6, 10 to 20, 23 to 27, 30, 31 and 32).

[0012] In one embodiment preferred nucleic acids encode at least two, more preferably three, still more preferably four, or most preferably or more ORFs selected from ORFS 1 to 32 (SEQ ID NOS: 2 to 33) of the ramoplanin locus. In one embodiment, combinations of ORFs selected from ORFs 1 through 32 (SEQ ID NOS 2 to 33) are provided which encode polypeptides that form at least the depsipeptide core structure of ramoplanin. In another embodiment combinations of ORFs selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33) are provided which encode polypeptides that form at least the fatty-aid side chain of the depsipeptide core structure of ramoplanin. In another embodiment, combinations of ORFs selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33) are provided which encode polypeptides responsible for the synthesis of 4-hydroxyphenylglycine (HPG) of ramoplanin. In another embodiment, combinations of ORFs selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33) are provided that encode polypeptides that form at least the beta-hydroxyasparagine residue. In another embodiment, combinations of ORFs selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33) are provided which are involved in the regulation of ramoplanin biosynthesis. In another embodiment, combinations of ORFs selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33) are provided which encode polypeptides that are involved in resistance and subcellular localization of the ramoplanin biosynthetic machinery. A single ORF or a combination of ORFs selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33) are provided to enhance production of ramoplanin by altering the expression level of an ORF selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33). In another embodiment, the expression level of an ORF selected from ORFs 1 through 32 (SEQ ID NOS: 2 to 33) may be altered to increase the yield of a particular form of ramoplanin.

[0013] Those skilled in the art will readily understand that the invention, having provided the polynucleotide sequences encoding polypeptides of the ramoplanin biosynthetic pathway, also provides polynucleotides encoding fragments derived from such peptides. Moreover, the invention is understood to provide naturally occurring variants or derivatives of such polypeptides and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids as described herein. Those skilled in the art would also readily understand that the invention, having provided the polynucleotide sequences of the entire genetic locus from *Actinoplanes*, further provides naturally-occurring variants or homologs of the

genes of the ramoplanin biosynthetic locus from other microorganisms, in particular, those of the family *Actinomycetes*.

[0014] It is also understood that the invention, having provided the polynucleotide sequences of the entire genetic locus as well as the coding sequences, further provides polynucleotides which regulate the expression of the polypeptides of the biosynthetic pathway. Such regulating polynucleotides include but are not limited to promoter and enhancer sequences, as well as sequences antisense to any of the aforementioned sequences. The antisense molecules are regulators of gene expression in that they are used to suppress expression of the gene from which they are derived. Expression cassettes and vectors comprising a polynucleotide as described herein, as well as cells transformed or transfected with such cassettes and vectors, are also within the scope of the invention.

[0015] In one aspect, the invention provides polynucleotides encoding a polypeptide selected from ORFs 9, 11 to 15, 17, 26 and 27 (SEQ ID NOS: 10, 12 to 16, 18, 27 and 28) or naturally occurring variants or derivatives of such polypeptides and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of any one of ORFs 9, 11 to 15, 17, 26 and 27, for use in the synthesis of ramoplanin *in vivo* or *in vitro*. Such polynucleotides and polypeptides may also be used to generate derivatives of ramoplanin. In one embodiment, the order in which the modules occur within a single ORF may be changed so that the ramoplanin core structure is altered. In another embodiment, one or more module from one or more ORFs may be deleted or inserted so that the size of the ramoplanin core is altered. The polynucleotides and polypeptides related to ORFs 9, 11 to 15, 17, 26 and 27 may also be used to improve production or to produce variants of other antibiotics of the peptide class. In one embodiment, a module contained in any one or more of ORFs 9, 11 to 15, 17, 26 and 27 may be used to replace an existing module in a peptide synthetase involved in the synthesis of another peptide antibiotic to produce a peptide antibiotic derivative. In another embodiment, a module contained in any one or more of ORFs 9, 11 to 15, 17, 26 and 27 may be inserted into the sequence encoding the peptide synthetase involved in the synthesis of another peptide antibiotic to produce a peptide antibiotic derivative with a longer peptide length. In another embodiment, a module contained in any one or more of ORFs 9, 11 to 15, 17, 26 and 27 may be used in combination with the sequences of the present invention or in combination with other

sequences which encode other peptide synthetases, to custom design a peptide antibiotic.

[0016] In another aspect, the invention provides polynucleotides encoding ORF17 (SEQ ID NOS: 18), or naturally occurring variants or derivatives of ORF17 and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of ORF17, for use as an adenylation domain in conjunction with other peptide synthetase modules and allowing the incorporation of Thr into a peptide antibiotic precursor.

[0017] In another aspect, the invention provides polynucleotides encoding ORF 11, 12 or 26 (SEQ ID NOS: 12, 13 and 27), or naturally occurring variants or derivatives of ORF11, 12 or 26 and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of ORF11, 12 or 26, for incorporating fatty acids into the core structure of a peptide antibiotic precursor. In one embodiment, ORF16, 24 or 25 or their variant or derivative is used in conjunction with ORF11, 12 or 26, for modifying fatty acid structure and/or enhancing fatty acid incorporation into the peptide antibiotic structure. In another embodiment, ORF1, 3, 19 or 29 or their variant or derivative is used in conjunction with ORF11, 12 or 26, for further enhancing fatty acid incorporation into the peptide antibiotic structure.

[0018] In another aspect, the invention provides polynucleotides encoding the adenylation and/or condensation domain of a module selected from module 1, 2, 3 and 5 of ORF 13 (SEQ ID NO: 14) and modules 1, 3 and 7 of ORF 14 (SEQ ID NO: 15), or naturally occurring variants or derivatives of such polypeptides and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of an adenylation domain of a module selected from modules 1, 2, 3 and 5 of ORF 13 (SEQ ID NO: 14) and modules 1, 3 and 7 of ORF 14, for incorporating a D-amino acid into the core structure of a peptide antibiotic precursor.

[0019] In another aspect, the invention provides polynucleotides encoding any one of ORFs 4, 6, 7, 28 and 30 (SEQ ID NOS: 5, 7, 8, 29 and 31), or naturally occurring variants or derivatives of ORFs 4, 6, 7, 28 or 30 and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of ORF 4,

6, 7, 28 or 30, for synthesis of hydroxyphenylglycine (HPG). In one embodiment, any one of ORFs 4, 6, 7, 28 and 30 or their variant or derivative is used to enhance production of an HPG-containing peptide antibiotic, including but not limited to nocardicin A, vancomycin, aridicin, chloroeremomycin, teicoplanin and related glycopeptide antibiotics, as well as the calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor.*

[0020] In another aspect, the invention provides polynucleotides encoding any one of ORFs 2, 3, 8, 19, 23, 29 and 31 (SEQ ID NOS: 3, 4, 9, 20, 24, 30 and 32), or naturally occurring variants or derivatives of ORF 2, 3, 8, 19, 23, 29 or 31 and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of ORF 2, 3, 8, 19, 23, 29 or 31, for enhancing secretion of ramoplanin or its variants and derivatives, or for enhancing uptake of precursors for ramoplanin biosynthesis. In one embodiment, any one of ORFs 2, 8, 23 and 31 may be used to confer resistance to ramoplanin or its variants and derivatives or improve production levels.

[0021] In another aspect, the invention provides polynucleotides encoding any one of ORFs 5, 21 and 22 (SEQ ID NOS: 6, 22 and 23), or naturally occurring variants or derivatives of ORF 5, 21 or 22 and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of ORF 5, 21 or 22, for regulating biosynthesis of ramoplanin or its variants and derivatives. In one embodiment, any one of ORFs 5, 21 and 22 may be used to enhance production of ramoplanin or its variants and derivatives. In another embodiment, any one of ORFs 5, 21 and 22 may be used to link expression of ramoplanin or its variants and derivatives to an environmental or cellular signal.

[0022] In another aspect, the invention provides polynucleotides encoding ORF20 (SEQ ID NO: 21), or naturally occurring variants or derivatives of ORF20 and fragments derived therefrom, such variants or derivatives resulting from the addition, deletion, or substitution of non-essential amino acids or conservative substitutions of essential amino acids of ORF20, for halogenation of aromatic groups of a peptide antibiotic precursor. In one embodiment, ORF20 or its variants or derivatives are used to chlorinate HPG of a peptide antibiotic precursor.

BRIEF DESCRIPTION OF THE DRAWINGS:

[0023] Various embodiments of the invention will now be described with reference to the attached Figures:

[0024] Figure 1 is a graphical depiction of the ramoplanin biosynthetic locus showing a scale in kb, the relative position and orientation of the 32 ORFs, and the coverage of the deposited cosmids.

[0025] Figure 2A is a model for the biosynthesis of ramoplanin. The ramoplanin chain is assembled in stepwise fashion through the concerted activities of consecutive modules of the ramoplanin peptide synthetases. Domains in each module are denoted by the circular and oval symbols as indicated. R denotes the fatty acyl group that caps the N-terminus of the first amino acid (Asn) incorporated into the ramoplanin peptide (see Figure 2B). Note that ORF 12 recognizes Asn and is proposed to incorporate both Asn residues found in the ramoplanin peptide; hydroxylation of the second Asn residue may occur before or after recognition and activation of the amino acid. The thick dotted arrow indicates that the ORF 17 protein interacts with module 6 of the ORF 13 product to catalyze the incorporation of Thr at the appropriate position. The thin dotted line indicates that the side chain hydroxyl group of the beta-hydroxyasparagine residue undergoes nucleophilic attack upon the thioester bond linking the ramoplanin product with module 8 of ORF 14, resulting in the cyclization and release of the peptide product. Abbreviations: HAsn, beta-hydroxyasparagine; other abbreviations are as in the text.

[0026] Figure 2B is a model for the initiation of ramoplanin peptide synthesis using a fatty acid starter group. ORF 11 and ORF 26 are proposed to act coordinately as a starter unit, using a fatty acid group to prime the assembly of the peptide chain. Symbols are as in Figure 2A.

[0027] Figure 2C illustrates the structure of ramoplanin. Shown are the positions of amino acid substituents, as well as an embodiment wherein the acylamide moiety is derived from an eight-carbon fatty acid (R). Alternative fatty acyl chaims may also be incorporated at this position.

[0028] Figure 3A is a clustal analysis of adenylation domains of ramoplanin biosynthetic enzymes. Shown is the alignment of the amino acid sequence (single letter code) of all adenylation domains found in the ramoplanin locus relative to the adenylation domain of gramicidin S synthetase GrsA. Adenylation domains of multimodular non-ribosomal peptide synthetases ORF13 and ORF14 are labeled

according to their corresponding module M1-M7 and M1-M8, respectively. Note that ORF13 does not contain an adenylation domain in module 6. Highly conserved core motifs A1-A10 of adenylation domains (Konz et al., 1999, Chem. Biol. Vol. 6, pp. R39-48) are highlighted by boxes. Key residues used to predict the substrate specificity of each adenylation domain are highlighted in black (see Figure 3B).

[0029] Figure 3B shows the predicted specificities of adenylation domains. The model of Challis et al. (Chem. Biol. 2000, Vol. 7, pp. 211-224) was used to extract key residues predicted to dictate the amino acid specificity of each adenylation domain (highlighted in black in Figure 3A). The corresponding eight residues that align with GrsA amino acids 235, 236, 239, 278, 299, 301, 322, and 330 are grouped with signatures of adenylation domains of known specificities (data kindly provided by Jacques Ravel). The accession number, protein name, and module number as well as the known amino acid specificity is shown for the latter. Abbreviations: Cda, CDA peptide synthetase of *Streptomyces coelicolor*; Cep, chloroeremomycin peptide synthetase of *Amycolatopsis orientalis*; Acm, actinomycin synthetase of *Streptomyces chrysomallus*; Fen, fengycin peptide synthetase of *Bacillus subtilis*; Bac, bacitracin peptide synthetase of *Bacillus licheniformis*; Fxb, exochelin peptide synthetase of *Mycobacterium smegmatis*; Tyc, tyrocidine peptide synthetase of *Brevibacillus brevis*; GrsA, gramicidin peptide synthetase of *Bacillus brevis*; DhbF, siderophore 2,3-dihydroxybenzoate synthetase of *Bacillus subtilis*; Nos, nostopeptolide peptide synthetase of Nostoc sp.; Css, cyclosporine peptide synthetase of *Tolypocladium inflatum*; HPG, 4-hydroxy-phenylglycine; 5hOrn, 5-hydroxyornithine; Pch, pyochelin of *Pseudomonas aeruginosa*.

[0030] Figure 3C shows the similarity between ORF26 and acyl-CoA ligases. Shown is the clustal analysis of ORF 26 versus several acyl-Coenzyme A ligases from diverse species: Mb, *Mycobacterium bovis*; Mt, *Mycobacterium tuberculosis*; Sv, *Streptomyces verticillus*; Mx, *Myxococcus xanthus*; Bs, *Bacillus subtilis*. Highlighted by boxes are the highly conserved core motifs AL1-AL8 of acyl-CoA ligases as described by Du et al., 2000.

[0031] Figure 4 illustrates the proposed biosynthetic pathway of the unusual amino acid 4-hydroxyphenylglycine (HPG). Chorismate (1), prephenate (2) and 4-hydroxyphenylpyruvate (3) are intermediates in the biosynthesis of the amino acid tyrosine (4). ORF 28 shows similarity to chorismate mutases of primary metabolism and therefore may catalyze the conversion of (1) to (2). ORF 4 shows amino acid

similarity to prephenate dehydrogenases of primary metabolism and therefore may catalyze the conversion of (2) to (3). ORF 30 shows amino acid similarity to 4-hydroxyphenylpyruvate dioxygenases, which convert (3) to homogentisate (5), an important intermediate in the metabolism of tyrosine. ORF30 may therefore catalyze a similar oxidative decarboxylation reaction to generate 4-hydroxymandelate (6). ORF 7 shows amino acid similarity to glycolate oxidases, which catalyze the conversion of glycolate to glyoxalate. ORF 7 may therefore convert the glycolate structure found in (6) to the corresponding glyoxalate structure to produce 4-hydroxyphenylglyoxalate (7). ORF 6 shows amino acid similarity to many aminotransferases, and may catalyze the conversion of (7) to HPG (8). Biochemical studies with radiolabeled amino acids have established that the HPG residues of the antibiotic vancomycin are derived from tyrosine, and structures 6, 7, and 8 were proposed as possible intermediates in HPG biosynthesis (Nicas et al., in Biotechnology of Antibiotics, Marcel Dekker, Inc., 1997, pp. 363-392 and references therein).

[0032] Figure 5 illustrates two clustal alignments. Figure 5A shows the local amino acid sequence homology between ORF 10 (SEQ ID NO: 11) and a key motif found in pfam 00753 involved in coordinating two zinc molecules in the beta-lactamase superfamily. (For information regarding the Pfam Families Datebase, see Bate *et al.* Nucleic Acids Rsearch, 2000, Vol. 28, No. 1). 1SML represents one member of this superfamily for which a crystal structure showing the intimate interaction between the zinc molecule and the highlighted residues is available (Ullah et al., J. Mol Biol., 1998 Nov 20; 284(1):125-36). Figure 5B shows the local amino acid sequence homology between ORF 10 (SEQ ID NO: 11) and a key motif found in pfam 00067 involved in coordinating an iron molecule in cytochrome P450 monooxygenases.

[0033] Figure 6 illustrates a RT-PCR analysis of recombinant *S. lividans* clones expressing ramoplanin ORF 10 (SEQ ID NO: 11).

[0034] Figure 7 illustrates a SDS-PAGE analysis of recombinant *S. lividans* clones expressing ramoplanin ORF 10 (SEQ ID NO: 11).

[0035] DETAILED DESCRIPTION OF THE INVENTION:

Ramoplanins are naturally produced by the microorganism *Actinoplanes sp.* ATCC 33076. The genetic locus encoding the biosynthetic pathway for ramoplanin production was isolated and cloned by the procedure described in USSN 09/910,813, from genomic DNA isolated from a ramoplanin producing strain of *Actinoplanes sp.*

ATCC 33076 (obtained from the American Type Culture Collection, Manassas, VA, USA). This newly discovered locus encodes 32 individual proteins involved in the biosynthesis of ramoplanin by this organism. The 32 proteins are encoded by ORFs contained within the contiguous sequence of 88421 base pairs of DNA (SEQ ID NO: 1).

[0036] Three deposits, namely *E. coli* DH10B (008CH) strain, *E. coli* DH10B (008CK) strain and *E. coli* DH10B (008CO) strain each harbouring a cosmid clone of a partial biosynthetic locus for ramoplanin have been deposited with the International Depositary Authority of Canada, Bureau of Microbiology, Health Canada, 1015 Arlington Street, Winnipeg, Manitoba, Canada, R3E 3R2 on September 19, 2001. Clone 008CH, which spans from base pair 5006 to base pair 42974 of SEQ ID NO: 1, was assigned accession number IDAC 190901-3. Clone 008CK, which spans from base pair 34296 to base pair 70934 of SEQ ID NO: 1, was assigned accession number IDAC 190901-1. Clone 008CO, which spans from base pair 52163 to base pair 88333 of SEQ ID NO: 1, was assigned accession number IDAC 190901-2. The cosmids deposited as *E. coli* strains harbouring them are referred to herein as "the deposited cosmids".

[0037] As shown in Figure 1, the deposited cosmids comprise the biosynthetic locus for ramoplanin. The sequence of the polynucleotides comprised in the deposited cosmids, as well as the amino acid sequence of any polypeptide encoded thereby are controlling in the event of any conflict with any description of sequences herein.

[0038] The deposit of the cosmids has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The deposited cosmids will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited cosmids are provided merely as convenience to those skilled in the art and are not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112. A license may be required to make, use or sell the deposited cosmids, and compounds derived therefrom, and no such license is hereby granted.

[0039] Various reagents of the inventions can be isolated from the deposited strains. DNA sequence analysis was performed on various subclones of the inventions and facilitated the identification of the location of various ramoplanin ORFs, including the ORFs encoding the 32 individual proteins of the ramoplanin biosynthetic locus.

[0040] The ramoplanin biosynthetic locus spans approximately 88,500 base pairs and contains 32 ORFs. The contiguous nucleotide sequence of SEQ ID NO: 1 (88421 base pairs) contains the 33 deduced proteins listed in SEQ ID NOS: 2 to 34. ORF 1 (SEQ ID

NO: 2) represents 333 amino acids deduced from residues 2077 to 3078 (sense strand) of SEQ ID NO: 1. ORF 2 (SEQ ID NO: 3) represents 304 amino acids deduced from residues 3118 to 4032 (sense strand) of SEQ ID NO: 1. ORF 3 (SEQ ID NO: 4) represents 336 amino acids deduced from residues 4038 to 5048 (sense strand) of SEQ ID NO: 1. ORF 4 (SEQ ID NO: 5) represents 283 amino acids deduced from residues 6665 to 5814 (antisense strand) of SEQ ID NO: 1. ORF 5 (SEQ ID NO: 6) represents 336 amino acids deduced from residues 7703 to 6693 (antisense strand) of SEQ ID NO: 1. ORF 6 (SEQ ID NO: 7) represents 444 amino acids deduced from residues 9464 to 8130 (antisense strand) of SEQ ID NO: 1. ORF 7 (SEQ ID NO: 8) represents 356 amino acids deduced from residues 9691 to 10761 (sense strand) of SEQ ID NO: 1. ORF 8 (SEQ ID NO: 9) represents 640 amino acids deduced from residues 12751 to 10829 (antisense strand) of SEQ ID NO: 1. ORF 9 (SEQ ID NO: 10) represents 271 amino acids deduced from residues 13617 to 12802 (antisense strand) of SEQ ID NO: 1. ORF 10 (SEQ ID NO: 11) represents 529 amino acids deduced from residues 15203 to 13614 (antisense strand) of SEQ ID NO: 1. ORF 11 (SEQ ID NO: 12) represents 90 amino acids deduced from residues 15591 to 15863 (sense strand) of SEQ ID NO: 1. ORF 12 (SEQ ID NO: 13) represents 1051 amino acids deduced from residues 15880 to 19035 (sense strand) of SEQ ID NO: 1. ORF 13 (SEQ ID NO: 14) represents 6893 amino acids deduced from residues 19032 to 39713 (sense strand) of SEQ ID NO: 1. ORF 14 (SEQ ID NO: 15) represents 8695 amino acids deduced from residues 39713 to 65800 (sense strand) of SEQ ID NO: 1. ORF 15 (SEQ ID NO: 16) represents 234 amino acids deduced from residues 65826 to 66530 (sense strand) of SEQ ID NO: 1. ORF 16 (SEQ ID NO: 17) represents 274 amino acids deduced from residues 66546 and 67370 (sense strand) of SEQ ID NO: 1. ORF 17 (SEQ ID NO: 18) represents 891 amino acids deduced from residues 67384 to 70059 (sense strand) of SEQ ID NO: 1. ORF 18 (SEQ ID NO: 19) represents 187 amino acids deduced from residues 70099 to 70662 (sense strand) of SEQ ID NO: 1. ORF 19 (SEQ ID NO: 20) represents 415 amino acids deduced from residues 70659 to 71906 (sense strand) of SEQ ID NO: 1. ORF 20 (SEQ ID NO: 21) represents 491 amino acids deduced from residues 73439 to 71964 (antisense strand) of SEQ ID NO: 1. ORF 21 (SEQ ID NO: 22) represents 217 amino acids deduced from residues74216 to 73563 (antisense strand) of SEQ ID NO: 1. ORF 22 (SEQ ID NO: 23) represents 403 amino acids deduced from residues 75424 to 74213 (antisense strand) of SEQ ID NO: 1. ORF 23 (SEQ ID NO: 24) represents 309 amino acids deduced from residues 75535 to

76464 (sense strand) of SEQ ID NO: 1. ORF 24 (SEQ ID NO: 25) represents 553 amino acids deduced from residues 78110 to 76449 (antisense strand) of SEQ ID NO: 1. ORF 25 (SEQ ID NO: 26) represents 585 amino acids deduced from residues 79864 to 78107 (antisense strand) of SEQ ID NO: 1. ORF 26 (SEQ ID NO: 27) represents 587 amino acids deduced from residues 81624 to 79861 (antisense strand) of SEQ ID NO: 1. ORF 27 (SEQ ID NO: 28) represents 75 amino acids deduced from residues 81909 to 81682 (antisense strand) of SEQ ID NO: 1. ORF 28 (SEQ ID NO: 29) represents 94 amino acids deduced from residues 82346 to 82062 (antisense strand) of SEQ ID NO: 1. ORF 29 (SEQ ID NO: 30) represents 619 amino acids deduced from residues 82587 to 84446 (sense strand) of SEQ ID NO: 1. ORF 30 (SEQ ID NO: 31) represents 355 amino acids deduced from residues 84481 to 85548 (sense strand) of SEQ ID NO: 1. ORF 31 (SEQ ID NO: 32) represents 429 amino acids deduced from residues 85556 to 86845 (sense strand) of SEQ ID NO: 1. ORF 32 (SEQ ID NO: 33) represents 189 amino acids deduced from residues 87372 to 86803 (antisense strand) of SEQ ID NO: 1. ORF 33 (SEQ ID NO: 34) is incomplete and represents 309 amino acids (N-terminus only) deduced from residues 87494 to 88420 (sense strand) of SEQ ID NO: 1.

[0041] Some ORFs, namely ORFs 4, 7, 8, 9, 12, 16, 17, 19, 20, 27, 28, 29, 30, 32, and 33 (SEQ ID NOS: 5, 8, 9, 10, 13, 17, 18, 20, 21, 25, 28, 29, 30, 31, 33 and 34) are initiated with the non-standard initiation codon GTG (Valine) rather than the standard initiation codon ATG (Methionine). All ORFs are listed with Methionine or Valine amino acids at the amino-terminal position to indicate the specificity of the first codon in the ORF. It is expected, however, that in all cases the biosynthesized protein will contain a methionine residue, and more specifically a formylmethionine residue, at the amino terminal position in keeping with widely accepted principle that protein synthesis in bacteria initiates with methionine (formylmethionine) even when the encoding gene specifies a non-standard initiation codon (see e.g. Stryer, Biochemistry 3rd edition, 1998, W.H. Freeman and Co., New York, pp. 752-754).

[0042] Section 1: Definitions

The term domain refers to a portion of a molecule, *e.g.* proteins or nucleic acids, that is structurally and/or functionally distinct from another portion of the molecule.

[0043] The term derivative or analog of a molecule refers to a portion derived from or a modified version of the molecule.

[0044] The term isolated nucleic acid molecule referred to in the present invention can be a deoxyribonucleic acid molecule (DNA), such as genomic DNA and complementary DNA (cDNA), which can be single (coding or noncoding strand) or double stranded, as well as synthetic DNA, such as synthesized, single stranded polynucleotide. The isolated nucleic acid molecule of the present invention can also be a ribonucleic acid molecule (RNA). In particular embodiments, the nucleic acid can include entire sequence of the gene cluster, the sequence of any one of the ORFs, a sequence encoding an ORF and an associated promoter, or smaller sequences useful for expressing peptides, polypeptides or full length proteins encoded in the fragment of the *Actinoplanes sp.* genome disclosed herein. In particular embodiments the nucleic acid can have natural, non-natural or modified nucleotides or internucleotide linkages or mixtures of these.

[0045] The term polynucleotide refers to full length or partial length sequences of ORFs disclosed herein. Polynucelotides of this invention can be either RNA or DNA (cDNA, genomic DNA or synthetic DNA), or modifications, variants, homologs or fragments thereof. If single stranded, the polynucleotides can be a coding or "sense" or positive strand or a complementary or "antisense" or negative strand. Antisense strands can be useful as modulators of the protein or proteins by interacting with RNA encoding the protein(s). Antisense strands are preferably less than full length strands having sequences unique or highly specific for RNA encoding the protein(s). Any one of the polynucleotide sequences of the invention as shown in the sequence listing is (a) a coding sequence, (b) a ribonucleotide sequence derived from transcription of (a), (c) a coding sequence which uses the redundancy or degeneracy of the genetic code to encode the same polypeptides, or (d) a regulatory sequence.

[0046] The term polypeptide or protein refers to any chain of amino acids, regardless of length or post-translational modification (*e.g.* proteolytic processing or phosphorylation). Both terms are used interchangeably in the present application. Those skilled in the art would readily understand that the polypeptides of the invention may be purified from a natural source, *i.e.*, an *Actinoplanes sp.*, or produced by recombinant means.

[0047] The terms ORF, ramoplanin open reading frame, and ramoplanin ORF refer to an open reading frame in the ramoplanin biosynthetic gene cluster as isolated from *Actinoplanes sp.* The term also embraces the same ORFs as present in other ramoplanin-synthesizing organisms (*e.g.* other strains and/or species of *Actinoplanes,*

Streptomyces, *Actinomycetes* , and the like). The term encompasses allelic variants and single nucleotide polymorphisms (SNPs). In certain instances the term ramoplanin ORF is used synonymously with the polypeptide encoded by the ramoplanin ORF and may include conservative substitutions in that polypeptide. The particular usage will be clear from context.

[0048] The term "homologous amino acid sequence" is any polypeptide which is encoded, in whole or in part, by a nucleic acid sequence which hybridizes at 25-35°C below critical melting temperature (Tm), to any portion of the coding region nucleic acid sequences of the sequence listing. A homologous amino acid sequence is one that differs from an amino acid sequence shown in the sequence listing by one or more conservative amino acid substitutions. Such a sequence also encompasses allelic variants (defined below) as well as sequences containing deletions or insertions which retain the functional characteristics of the polypeptide. Preferably, such a sequence is at least 75%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, and most preferably 98% identical to any amino acid sequence shown in the sequence listing.

[0049] Homologous amino acid sequences include sequences that are identical or substantially identical to the amino acid sequences of the sequence listing. By "amino acid sequence substantially identical" it is meant a sequence that is at least 90%, preferably 95%, more preferably 97%, and most preferably 99% identical to an amino acid sequence of reference and that preferably differs from the sequence of reference by a majority of conservative amino acid substitutions. Consistent with this aspect of the invention, polypeptides having a sequence homologous to any one of the amino acid sequences of the sequence listing include naturally-occurring allelic variants, as well as mutants or any other non-naturally occurring variants that retain the inherent characteristics of any polypeptide of the sequence listing.

[0050] Homology is measured using sequence analysis software such as Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705. Amino acid sequences are aligned to maximize identity. Gaps may be artificially introduced into the sequence to attain optimal alignment. Once the optimal alignment has been set up, the degree of homology is established by recording all of the positions in which the amino acids of both sequences are identical, relative to the total number of positions.

[0051] Homologous polynucleotide sequences are defined in a similar way. Preferably, a homologous sequence is one that is at least 45%, more preferably 60%, more preferably 75% and most preferably 85% identical to any one of the coding sequences of the sequence listing.

[0052] The term "conservative substitution" is used in reference to proteins or peptides to reflect amino acid substitutions that do not substantially alter the activity (specificity or binding affinity) of the molecule. Typically conservative amino acid substitutions involve substitutions of one amino acid for another amino acid with similar chemical properties (e.g. charge or hydrophobicity). The following six groups each contain amino acids that are typical conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0053] The terms "isolated", "purified", or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. With respect to nucleic acids and/or polypeptides, the term can refer to nucleic acids or polypeptides that are no longer flanked by the sequences typically flanking them in nature. Such isolated nucleic acids and/or polynucelotides may be part of a vector or composition and still be defined as isolated in that such a vector or composition is not part of the natural environment of such polynucleotide.

[0054] The term "heterologous" as it relates to nucleic acid sequences such as coding sequences and control sequences, denotes sequences that are not normally associated with a region of a recombinant construct, and/or are not normally associated with a particular cell. Thus, a "heterologous" region of a nucleic acid construct is an identifiable segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with the other molecule in nature. For example, a heterologous region of a construct could include a coding sequence flanked by sequences not found in association with the coding sequence in nature. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g. synthetic sequences having codons different than the native gene). Similarly, a host cell transformed with a construct which is not normally present in the host cell would be considered heterologous for purposes of this invention.

[0055] The term allelic variant refers to an alternate form of a polypeptide that is characterized as having a substitution, deletion, or addition of one or more amino acids that does not alter the biological function of the polypeptide.

[0056] The term "biological function" refers to the function of the polypeptide in the cells in which it naturally occurs. A polypeptide can have more than one biological function.

[0057] Section 2: Isolation, preparation and expression of ramoplanin nucleic acids

Nucleic acids derived from the ramoplanin gene cluster can be isolated, optionally modified and inserted into a host cell to create and/or modify a metabolic (biosynthetic) pathway and thereby enable that host cell to synthesize and/or modify various metabolites. Alternatively, the ramoplanin gene cluster nucleic acids can be expressed in the host cell and the encoded ramoplanin polypeptide(s) recovered for use as chemical reagents, *e.g.* in the *ex vivo* synthesis and/or chemical modification of various metabolites. Either application typically entails insertion of one or more nucleic acids encoding one or more isolated and/or modified ramoplanin ORFs in a suitable host cell. The nucleic acid(s) are typically in an expression vector, a construct containing control elements suitable to direct expression of the ramoplanin polypeptides. The expressed ramoplanin polypeptides in the host cell then act as components of a metabolic/biosynthetic pathway (in which case the synthetic product of the pathway is typically recovered) or the ramoplanin polypeptides themselves are recovered. Using the sequence information provided herein, cloning and expression of ramoplanin nucleic acids can be accomplished using routine and well known methods.

[0058] A. Ramoplanin nucleic acids

The nucleic acids comprising the ramoplanin gene cluster are identified in Table 2 and are listed in the sequence listing provided herein. In particular, Table 2 identifies genes and functions of ORFs in the ramoplanin biosynthetic gene cluster. Using the sequence information provided therein, primers suitable for amplification/isolation of one or more ORFs can be determined according to standard methods well known to those of skill in the art (*e.g.* using methods described in Innis (1990) *PCR Protocols: A Guide to Methods and Applications* Academic Press Inc. San Diego, CA, *etc*; using computer applications such as Vector NTI Suite™, InforMax, Gaithersberg, MD, USA).

[0059] Primers suitable for amplification/isolation of any one or more of the ORFs are designed according to the nucleotide sequence information provided in the sequence listing. The procedure is as follows: a primer is selected which consists of 10 to 40, preferably 15 to 25 nucleotides. It is advantageous to select primers containing C and G nucleotides in a proportion sufficient to ensure efficient hybridization; *i.e.*, an amount of C and G nucleotides of at least 40%, preferably 50% of the total nucleotide content. Typically such amplifications will utilize the DNA or RNA of an organism containing the requisite genes (*e.g. Actinoplanes sp.*) as a template. A standard PCR reaction contains typically 0.5 to 5 Units of Taq DNA polymerase per 100 μL, 20 to 200 μM deoxynucleotide each, preferably at equivalent concentrations, 0.5 to 2.5 mM magnesium over the total deoxynucleotide concentration, 10^5 to 10^6 target molecules, and about 20 pmol of each primer. About 25 to 50 PCR cycles are performed, with an annealing temperature 15°C to 5°C below the true Tm of the primers. A more stringent annealing temperature improves discrimination against incorrectly annealed primers and reduces incorportion of incorrect nucleotides at the 3' end of primers. A denaturation temperature of 95°C to 97°C is typical, although higher temperatures may be appropriate for denaturation of G+C-rich targets. Adding DMSO to a final concentration of 5-10% is beneficial for PCR amplification of high G+C templates such as those from *Actinoplanes sp.* The number of cycles performed depends on the starting concentration of target molecules, though typically more than 40 cycles is not recommended as non-specific background products tend to accumulate.

[0060] An alternative method for retrieving polynucleotides encoding homologous polypeptides or allelic variants is by hybridization screening of a DNA or RNA library. Hybridization procedures are well-known in the art and are described in Ausubel *et al.*, (Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons Inc., 1994), Silhavy *et al.* (Silhavy *et al.* Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, 1984), and Davis *et al.* (Davis *et al.* A Manual for Genetic Engineering: Advanced Bacterial Genetics, Cold Spring Harbor Laboratory Press, 1980)). Important parameters for optimizing hybridization conditions are reflected in a formula used to obtain the critical melting temperature above which two complementary DNA strands separate from each other (Casey & Davidson, Nucl. Acid Res. (1977) 4:1539). For polynucleotides of about 600 nucleotides or larger, this formula is as follows: Tm = 81.5 + 0.5 x (% G+C) + 1.6 log (positive ion concentration) - 0.6 x (%

formamide). Under appropriate stringency conditions, hybridization temperature (Th) is approximately 20 to 40°C, 20 to 25°C, or, preferably 30 to 40°C below the calculated Tm. Those skilled in the art will understand that optimal temperature and salt conditions can be readily determined.

[0061] For the polynucleotides of the invention, stringent conditions are achieved for both pre-hybridizing and hybridizing incubations (i) within 4-16 hours at 42°C, in 6x SSC containing 50% formamide, or (ii) within 4-16 hours at 65°C in an aqueous 6x SSC solution (1 M NaCl, 0.1M sodium citrate (pH 7.0)).

[0062] In one embodiment, this invention provides nucleic acids for the recombinant expression of a ramoplanin (*e.g.* a ramoplanin or an analogue thereof). Such nucleic acids include isolated gene cluster(s) comprising ORFs encoding polypeptides sufficient to direct the synthesis of the ramoplanin. In other embodiments of this invention, the ORFs may be unchanged, but the control elements (*e.g.* promoters, ribosome binding sites, terminators, enhancers *etc*) may be modified. In still other embodiments, the nucleic acids may encode selected components (*e.g.* one or more ORFs or modified ORFs) and/or may optionally contain other heterologous biosynthetic elements including, but not limited to non-ribosomal polypeptide synthetases (NRPS) modules or enzymatic domains.

[0063] Such variations may be introduced by design, for example to modify a known molecule in a specific way, *e.g.* by replacing a single substitutent of the ramoplanin with another, thereby creating a derivative ramoplanin molecule of predicted structure. Alternatively, variations can be made randomly, for example by making a library of molecular variants of a known ramoplanin by systematically or haphazardly replacing one or more ORFs in the biosynthetic pathway.

[0064] Useful homologs and fragments thereof that do not occur naturally are designed using known methods for identifying regions of a polypeptide that are likely to tolerate amino acid sequence changes and/or deletions. As an example, homologous polypeptides from different species are compared; conserved sequences are identified. The more divergent sequences are the most likely to tolerate sequence changes. Homology among sequences may be analyzed using the BLAST homology searching algorithm of Altschul *et al.*, Nucleic Acids Res.25:3389-3402 (1997).

[0065] Alternatively, identification of homologous polypeptides or polypeptide derivatives encoded by polynucleotides of the invention which have activity in the ramoplanin biosynthetic pathway may be achieved by screening for cross-reactivity with

an antibody raised against the polypeptide of reference having an amino acid sequence of SEQ ID NOS 2 to 34. The procedure is as follows: an antibody is raised against a purified reference polypeptide, a fusion polypeptide (for example, an expression product of MBP, GST, or His-tag systems), or a synthetic peptide derived from the reference polypeptide. Where an antibody is raised against a fusion polypeptide, two different fusion systems are employed. Specific antigenicity can be determined according to a number of methods, including Western blot (Towbin *et al.*, Proc. Natl. Acad. Sci. USA (1979) 76:4350), dot blot, and ELISA, as described below.

[0066] In a Western blot assay, the product to be screened, either as a purified preparation or a total *E. coli* extract, is submitted to SDS-Page electrophoresis as described by Laemmli (Nature (1970) 227:680). After transfer to a nitrocellulose membrane, the material is further incubated with the antibody diluted in the range of dilutions from about 1:5 to about 1:5000, preferably from about 1:100 to about 1:500. Specific antigenicity is shown once a band corresponding to the product exhibits reactivity at any of the dilutions in the above range.

[0067] In an ELISA assay, the product to be screened is preferably used as the coating antigen. A purified preparation is preferred, although a whole cell extract can also be used. Briefly, about 100 μl of a preparation at about 10 μg protein/ml are distributed into wells of a 96-well polycarbonate ELISA plate. The plate is incubated for 2 hours at 37°C then overnight at 4°C. The plate is washed with phosphate buffer saline (PBS) containing 0.05% Tween 20 (PBS/Tween buffer). The wells are saturated with 250 μl PBS containing 1% bovine serum albumin (BSA) to prevent non-specific antibody binding. After 1 hour incubation at 37°C, the plate is washed with PBS/Tween buffer. The antibody is serially diluted in PBS/Tween buffer containing 0.5% BSA. 100 μl of dilutions are added per well. The plate is incubated for 90 minutes at 37°C, washed and evaluated according to standard procedures. For example, a goat anti-rabbit peroxidase conjugate is added to the wells when specific antibodies were raised in rabbits. Incubation is carried out for 90 minutes at 37°C and the plate is washed. The reaction is developed with the appropriate substrate and the reaction is measured by colorimetry (absorbance measured spectrophotometrically). Under the above experimental conditions, a positive reaction is shown by O.D. values greater than a non immune control serum.

[0068] In a dot blot assay, a purified product is preferred, although a whole cell extract can also be used. Briefly, a solution of the product at about 100 μg/ml is serially

two-fold diluted in 50 mM Tris-HCl (pH 7.5). 100 μl of each dilution are applied to a nitrocellulose membrane 0.45 μm set in a 96-well dot blot apparatus (Biorad). The buffer is removed by applying vacuum to the system. Wells are washed by addition of 50 mM Tris-HCl (pH 7.5) and the membrane is air-dried. The membrane is saturated in blocking buffer (50 mM Tris-HCl (pH 7.5) 0.15 M NaCl, 10 g/L skim milk) and incubated with an antibody dilution from about 1:50 to about 1:5000, preferably about 1:500. The reaction is revealed according to standard procedures. For example, a goat anti-rabbit peroxidase conjugate is added to the wells when rabbit antibodies are used. Incubation is carried out 90 minutes at 37°C and the blot is washed. The reaction is developed with the appropriate substrate and stopped. The reaction is measured visually by the appearance of a colored spot, *e.g.*, by colorimetry. Under the above experimental conditions, a positive reaction is shown once a colored spot is associated with a dilution of at least about 1:5, preferably of at least about 1:500.

[0069] Using the information provided herein other approaches to cloning the desired sequences will be apparent to those of skill in the art, for example, the ramoplanin genes and/or optionally NRPS modules or enzymatic domains of interest can be obtained from an organism that expresses such, using recombinant methods, such as by screening cDNA or genomic libraries, derived from cells expressing the gene, or by deriving the gene from a vector known to include the same. The gene can then be isolated and combined with other desired biosynthetic elements using standard techniques. If the gene in question is already present in a suitable expression vector, it can be combined *in situ* with, *e.g.* other domains or subunits, as desired. The gene of interest can be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence can be assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence (*see e.g., Edge (1981) Nature* 292:756; Nambair *et al.* (1984) *Science* 233:1299; Jay *et al.* (1984) *J. Biol. Chem.* 259:6311). In addition, it is noted that custom gene synthesis is commercially available (*see e.g.* Operon Technologies, Alameda, CA).

[0070] Examples of such techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel (1989) *Guide to Molecular Cloning Technique, Methods in Enzymology 152* Academic Press, Inc., San

Diego, CA (Berger); Sambrook *et al.* (1989) *Molecular Cloning – A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, N.Y.; Ausubel (1994) *Current Protocols in Molecular Biology*, Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. U.S. Patent 5,017,478; and European Patent No 0 246 864.

[0071] B. Expression of ramoplanin ORFs

Preferably, a recombinant expression system is selected from prokaryotic hosts. Bacterial cells are available from a number of different sources including commercial sources to those skilled in the art, *e.g.*, the American Type Culture Collection (ATCC; Rockville, Maryland). Commercial sources of cells used for recombinant protein expression also provide instructions for usage of the cells.

[0072] The choice of the expression system depends on the features desired for the expressed polypeptide. For example, it may be useful to produce a polypeptide of the invention in a particular lipidated form or any other form. Any transducible cloning vector can be used as a cloning vector for the nucleic acid constructs of this invention. However, where large clusters are to be expressed, it is preferable that phagemids, cosmids, P1s, YACs, BACs, PACs, HACc or similar cloning vectors be used for cloning the nucleotide sequences into the host cell. Phagemids, cosmids, and BACs, for example, are advantageous vectors due to the ability to insert and stably propagate therein larger fragments of DNA than in M13 phage and lambda phage, respectively. Phagemids which will find use in this method generally include hybrids between plasmids and filamentous phage cloning vehicles. Cosmids which will find use in this method generally include lambda phage-based vectors into which cos sites have been inserted. Recipient pool cloning vectors can be any suitable plasmid. The cloning vectors into which pools of mutants are inserted may be identical or may be constructed to harbor and express different genetic markers (*see, e.g.,* Sambrook *et al., supra*). The utility of employing such vectors having different marker genes may be exploited to facilitate a determination of successful transduction.

[0073] In preferred embodiments of this invention, vectors are used to introduce ramoplanin biosynthesis genes or gene clusters into host (*e.g. Streptomyces*) cells. With the guidelines described below, however, a selection of vectors, expression control sequences and hosts may be made without undue experimentation and without departing from the scope of this invention. Numerous vectors for use in particular host

cells are well known to those of skill in the art. For example Malpartida and Hopwood, (1984) *Nature*, 309:462-464; Kao *et al.,* (1994), *Science,* 265: 509-512; and Hopwood *et al.,* (1987) *Methods Enzymol.*, 153:116-166 all describe vectors for use in various *Streptomyces* hosts. In selecting a vector, the appropriate host must be chosen such that it is compatible with the vector which is to exist and possibly replicate in it. Considerations are made with respect to the vector copy number, the ability to control the copy number and expression of other proteins such as antibiotic resistance. In one preferred embodiment, *Streptomyces* vectors are used that include sequences that allow their introduction and maintenance in *E. coli.* Such *Streptomyces/E. coli* shuttle vectors have been described (see, for example, Vara *et al.*, (1989) *J. Bacteriol*, 171:5872-5881; Guilfoile & Hutchinson (1991) *Proc. Natl. Acad. Sci. USA,* 88; 8553-8557.)

[0074] The wildtype and/or modified ORFs of this invention can be inserted into one or more expression vectors, using methods known to those of skill in the art. Expression vectors (*e.g.*, plasmids) are widely known and are readily available to those skilled in the art. For bacterial vectors, the polynucleotide of the invention is inserted into the bacterial genome or remains in a free state as part of a plasmid. Methods for transforming host cells with expression vectors are well-known in the art. Expression vectors will include control sequences operably linked to the desired ORF. In selecting an expression control sequence, a number of variables are considered. Among the important variables are the relative strength of the sequence (*e.g.* the ability to drive expression under various conditions), the ability to control the sequence's function and compatibility between the polynucleotide to be expressed and the control sequence (*e.g.* secondary structures are considered in order to avoid hairpin structures which may prevent efficient transcription).

[0075] Suitable expression systems for use with the present invention include systems that function in eucaryotic and/or prokaryotic host cells. However, as explained above, prokaryotic systems are preferred, and in particular, systems compatible with *Streptomyces sp.* are of particular interest.

[0076] The choice of the expression cassette depends on the host system selected as well as the features desired for the expressed polypeptide or natural product. Typically, an expression cassette includes a promoter that is functional in the selected host system and can be constitutive or inducible; a ribosome binding site; a start codon (ATG) if necessary; optionally a region encoding a leader peptide; a DNA molecule of

the invention; a stop codon; and optionally a 3' terminal region (translation and/or transcription terminator). Where applicable, *i.e.* secreted or membrane proteins, the leader peptide encoding region is adjacent to the polynucleotide of the invention and placed in proper reading frame. The leader peptide-encoding region, if present, is homologous or heterologous to the DNA molecule encoding the mature polypeptide and is compatible with the secretion apparatus of the host used for expression. The ORF constituted by the DNA molecule of the invention, solely or together with the leader peptide, is placed under the control of the promoter so that transcription and translation occur in the host system. Promoters and leader peptide encoding regions are widely known and available to those skilled in the art. Particularly useful promoters include control sequences derived from ramoplanin and/or NRPS gene clusters. Other bacterial promoters, such as those derived from sugar metabolizing enzymes, such as galactose, lactose *(lac)* and maltose, will also find use in the present constructs. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan *(trp),* the beta-lactamase *(bla)* promoter system, bacteriophase lambda PL, and T5. In addition, synthetic promoters (U.S. Patent 4,551,433), which do not occur in nature also function in bacterial host cells. In *Streptomyces*, numerous promoters have been described including constitutive promoters, such as *ErmE* and *TcmG* (Shen and Hutchinson, (1994) *J. Biol. Chem.* 269: 30726-30733), as well as controllable promoters such as *actI* and *actIII* (Pleper *et al.,* (1995) *Nature*, vol. 378: 263-266; Pieper *et al.,* (1995) *J. Am. Chem. Soc.*, 117: 11373-11374; and Wiesmann *et al.,* (1995) *Chem. & Biol.* 2: 583-589).

[0077] Other regulatory sequences may also be desirable which allow for regulation of expression of the ORFs relative to the growth of the host cell. Regulatory sequences are known to those skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other type of regulatory elements may also be present in the vector, for example, enhancer sequences.

[0078] Selectable markers can also be included in the recombinant expression vectors. A variety of markers are known which are useful in selecting for transformed cell lines and generally comprise a gene whose expression confers a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium. Such markers include, for example, genes that confer antibiotic resistance or sensitivity to the plasmid.

[0079] Various ramoplanin ORFs, and/or NRPS clusters or subunits of interest can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements, or under the control of, *e.g.,* a single promoter. The ORFs can include flanking restriction sites to allow for the easy deletion and insertion of other open reading frames so that hybrid synthetic pathways can be generated. The design of such unique restriction sites is known to those of skill in the art and can be accomplished using the techniques described above, such a site-directed mutagenesis and PCR.

[0080] Methods of cloning and expressing large nucleic acids such as gene clusters, including NRPS-encoding gene clusters, in cells including *Streptomyces* are well known to those skilled in the art (*see, e.g.,* Stutzman-Engwall and Hutchinson (1989) *Proc. Ntl. Acad. Sci. USA,* 86: 3135-3139: Motamedi and Hutchinson (1987) *Proc. Natl. Acad. Sci. USA,* 84: 4445-4449; Grim *et al.* (1994) *Gene,* 151 : 1-10; Kao *et al.* (1994) *Science,* 265 : 509-512; and Hopwood *et al.* (1987) *Meth. Enzymol.,* 153: 116-166). In some examples, nucleic acid sequences of well over 100 kb have been introduced into cells, including prokaryotic cells, using vector-based methods (*see* for example, Osoegawa *et al.,* (1998) *Genomics,* 52: 1-8; Woon *et al.*, (1996) *Nucl. Acids, Res.*, 24: 4202-4209).

[0081] C. Host cells

The vectors described above can be used to express various protein components of the ramoplanin and/or ramoplanin shunt metabolites, and/or other modified metabolites for subsequent isolation and/or to provide a biological synthesis of one or more desired biomolecules (*e.g.* ramoplanin and/or a ramoplanin analogue, etc). Where one or more proteins of the ramoplanin biosynthetic gene cluster are expressed (*e.g.* overexpressed) for subsequent isolation and/or characterization, the proteins are expressed in any prokaryotic or eukaryotic cell suitable for protein expression. In selecting the host, unicellular hosts are selected which are compatible with the selected vector, tolerant of any possible toxic effects of the expressed product, able to secrete the expressed product efficiently if such is desired, able to express the product in the desired conformation, easily scaled up, and having regard to ease of purification of the final product, which may be the expressed polypeptide or the natural product, e.g. an antibiotic, which is a product of the biosynthetic pathway of which the expressed

polypeptide is a part. In one preferred embodiment, the proteins are expressed in *E. coli.*

[0082] Host cells for the recombinant production of the ramoplanin, ramoplanin metabolites, shunt metabolites, etc. can be derived from any organism with the capability of harboring a recombinant ramoplanin gene cluster and/or subset thereof. Thus, the host cells of the present invention can be derived from either prokaryotic or eucaryotic organisms. Preferred host cells are those of species or strains (*e.g.* bacterial strains) that naturally express ramoplanin. Suitable host cells include, but are not limited to *Actinomycetes, Actinoplanetes,* and *Streptomycetes, Actinomadura, Micromonospra,* and the like. Particularly preferred host cells include, but are not limited to *Streptomyces globisporus, Streptomyces lividans, Streptomyces coelicolor, Microsmonospora echinospora* spp. *calichenisis, Actionamadura verrucosopora, Micromonospora chersina,* and *Streptomyces carzinostaticus.*

[0083] D. Recovery of the expression product

Recovery of the expression product (e.g., ramoplanin, ramoplanin analog, ramoplanin biosynthetic pathway polypeptide, etc.) is accomplished according to standard methods well known to those skilled in the art. Thus for example where ramoplanin biosynthetic gene cluster proteins are to be expressed and isolated, the proteins can be expressed with a convenient tag to facilitate isolation (*e.g.* a His_6) tag. Other standard protein purification techniques are suitable and well known to those of skill in the art (*see, e.g.* (Quadri *et al.* 1998) *Biochemistry* 37: 1585-1595; Nakano *et al.* (1992) *Mol. Gen. Genet.* 232: 313-321, *etc*).

[0084] A polypeptide or polypeptide derivative of the invention may be purified by affinity chromatography using as a ligand either an antibody or a compound related to ramoplanin or other lipodepsipeptide which binds to the polypeptide. The antibody is either polyclonal or monoclonal. Purified IgGs are prepared from an antiserum using standard methods (see, *e.g.*, Coligan *et al.*, Current Protocols in Immunology (1994) John Wiley & Sons, Inc., New York, NY). Conventional chromatography supports are described in, *e.g.*, Antibodies: A Laboratory Manual, D. Lane, E. Harlow, Eds. (1988).

[0085] Consistent with this aspect of the invention, polypeptide derivatives are provided that are partial sequences of the amino acid sequences of SEQ ID NOS: 2 to 34, partial sequences of polypeptide sequences homologous to the amino acid

sequences of SEQ ID NOS: 2 to 34, polypeptides derived from full-length polypeptides by internal deletion, and fusion proteins.

[0086] Polynucleotides encoding polypeptide fragments and polypeptides having large internal deletions are constructed using standard methods (Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons Inc., 1994). Such methods include standard PCR, inverse PCR, restriction enzyme treatment of cloned DNA molecules, or the method of Kunkel *et al.* (Kunkel *et al.* Proc. Natl. Acad. Sci. USA (1985) 82:448). Components for these methods and instructions for their use are readily available from various commercial sources such as Stratagene. Once the deletion mutants have been constructed, they are tested for their ability to improve production of ramoplanin or generate novel analogues of the antibiotic or natural products of the lipodepsipeptide class as described herein.

[0087] A fusion polypeptide is one that contains a polypeptide or a polypeptide derivative of the invention fused at the N- or C-terminal end to any other polypeptide (hereinafter referred to as a peptide tail). A simple way to obtain such a fusion polypeptide is by translation of an in-frame fusion of the polynucleotide sequences, *i.e.*, a hybrid gene. The hybrid gene encoding the fusion polypeptide is inserted into an expression vector which is used to transform or transfect a host cell. Alternatively, the polynucleotide sequence encoding the polypeptide or polypeptide derivative is inserted into an expression vector in which the polynucleotide encoding the peptide tail is already present. Such vectors and instructions for their use are commercially available, *e.g.* the pMal-c2 or pMal-p2 system from New England Biolabs, in which the peptide tail is a maltose binding protein, the glutathione-S-transferase system of Pharmacia, or the His-Tag system available from Novagen. These and other expression systems provide convenient means for further purification of polypeptides and derivatives of the invention.

[0088] Polynucleotides of 30 to 600 nucleotides encoding partial sequences of sequences homologous to nucleotide sequences of SEQ ID NOS: 2 to 34 are retrieved by PCR amplification using the parameters outlined above and using primers matching the sequences upstream and downstream of the 5' and 3' ends of the fragment to be amplified. The template polynucleotide for such amplification is either the full length polynucleotide homologous to a polynucleotide sequence of SEQ ID NOS: 2 to 34, or a polynucleotide contained in a mixture of polynucleotides such as a DNA or RNA library. As an alternative method for retrieving the partial sequences, screening hybridization is

carried out under conditions described above and using the formula for calculating Tm. Where fragments of 30 to 600 nucleotides are to be retrieved, the calculated Tm is corrected by subtracting (600/polynucleotide size in base pairs) and the stringency conditions are defined by a hybridization temperature that is 5 to 10°C below Tm. Where oligonucleotides shorter than 20-30 bases are to be obtained, the formula for calculating the Tm is as follows: Tm = 4 x (G+C) + 2 x (A+T). For example, an 18 nucleotide fragment of 50% G+C would have an approximate Tm of 54°C. Short peptides that are fragments of the polypeptide sequences of SEQ IS NOS: 2 to 34 or their homologous sequences, are obtained directly by chemical synthesis (E. Gross and H. J. Meinhofer, 4 The Peptides: Analysis, Synthesis, Biology; Modern Techniques of Peptide Synthesis, John Wiley & Sons (1981), and M. Bodanzki, Principles of Peptide Synthesis, Springer –Verlag (1984)).

[0089] Where components (*e.g.* ramoplanin ORFs) are used to synthesize and/or modify various biomolecules (*e.g.*ramoplanins, ramoplanin analogues, shunt metabolites, or even compounds unrelated to ramoplanin, *i.e.* biocatalysts) the desired product and/or shunt metabolites(s) are isolated according to standard methods well known to those of skill in the art (*see,. e.g.,* Carreras and Khosla (1998) *Biochemistry* 37: 2084-2088, Deutscher (1990) *Methods in Ensymology Volume 182: Guide to Protein Purification,* M. Deutscher, ed.

[0090] E. Probes

The sequence information provided in the present application enables the design of specific nucleotide probes and primers that are used for identifying and isolating putative lipdepsipeptide-producing microorganisms. Accordingly, an aspect of the invention provides a nucleotide probe or primer having a sequence found in or derived by degeneracy of the genetic code from a sequence shown in the sequence listing.

[0091] The term "probe" as used in the present application refers to DNA (preferably single stranded) or RNA molecules (or modifications or combinations thereof) that hybridize under the stringent conditions, as defined above, to nucleic acid molecules of SEQ ID NOS: 1 to 34, or to sequences homologous to those of SEQ ID NOS: 1 to 34, or to their complementary or anti-sense sequences. Generally, probes are significantly shorter than full-length sequences. Such probes contain from about 5 to about 100, preferably from about 10 to about 80, nucleotides. In particular, probes have sequences that are at least 75%, preferably at least 85%, more preferably 95%

homologous to a portion of a sequence disclosed in SEQ ID NOS: 1 to 34 or that are complementary to such sequences. Probes may contain modified bases such as inosine, methyl-5-deoxycytidine, deoxyuridine, dimethylamino-5-deoxyuridine, or diamino-2, 6-purine. Sugar or phosphate residues may also be modified or substituted. For example, a deoxyribose residue may be replaced by a polyamide (Nielsen *et al.*, Science (1991) 254:1497) and phosphate residues may be replaced by ester groups such as diphosphate, alkyl, arylphosphonate and phosphorothioate esters. In addition, the 2'-hydroxyl group on ribonucleotides may be modified by including such groups as alkyl groups.

[0092] Probes of the invention are used for identifying and isolating putative lipdepsipeptide-producing microorganisms, as capture or detection probes. Such capture probes are conventionally immobilized on a solid support, directly or indirectly, by covalent means or by passive adsorption. A detection probe is labeled by a detection marker selected from: radioactive isotopes, enzymes such as peroxidase, alkaline phosphatase, enzymes able to hydrolyze a chromogenic or fluorogenic or luminescent substrate, compounds that are chromogenic or fluorogenic or luminescent, nucleotide base analogs, and biotin.

[0093] Probes of the invention are used in any conventional hybridization technique, such as dot blot (Maniatis *et al.*, Molecular Cloning: A Laboratory Manual (1982) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), Southern blot (Southern, J. Mol. Biol. (1975) 98:503), northern blot (identical to Southern blot with the exception that RNA is used as a target), or the sandwich technique (Dunn *et al.*, Cell (1977) 12:23). The latter technique involves the use of a specific capture probe and/or a specific detection probe with nucleotide sequences that at least partially differ from each other.

[0094] A primer is usually about 10 to about 40 nucleotides that is used to initiate enzymatic polymerization of DNA in an amplification process (*e.g.*, PCR), in an elongation process, or in a reverse transcription method. Primers used in diagnostic methods involving PCR are labeled by methods known in the art. Primers can also be used as probes.

[0095] As described herein, the invention also encompasses (i) a reagent comprising a probe of the invention for detecting and/or isolating putative lipdepsipeptide-producing microorganisms; (ii) a method for detecting and/or isolating putative lipdepsipeptide-producing microorganisms, in which DNA or RNA is extracted from the microorganism

and denatured, and exposed to a probe of the invention, for example, a capture probe or detection probe or both, under stringent hybridization conditions, such that hybridization is detected; and (iii) a method for detecting and/or isolating putative lipdepsipeptide-producing microorganisms, in which (a) a sample is recovered or derived from the microorganism, (b) DNA is extracted therefrom, (c) the extracted DNA is primed with at least one, and preferably two, primers of the invention and amplified by polymerase chain reaction, and (d) the amplified DNA fragment is produced.

[0096] Examples: The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1: Identification of the ramoplanin biosynthetic locus in *Actinoplanes sp.* ATCC 33076.

Actinoplanes sp. ATCC 33076 was previously shown to naturally produce ramoplanins, a group of biologically active lipodepsipeptides (U.S. Patent No. 4,303,646). The genetic locus involved in the production of this compound was not previously identified. *Actinoplanes sp.* ATCC 33076 was obtained from the American Tissue Culture Collection (ATCC) Manassas, VA, and cultured according to standard microbiological techniques (Kieser et al. *Practical Streptomyces Genetics*, John Innes Centre, Norwich Research Part, Colney, Norwich NR4 7UH, England, 2000). Confluent mycelia from oatmeal agar plates were used for the extraction of genomic DNA as previously described (Kieser et al., *supra*) and the size range of the DNA obtained was assessed on agarose gels by electrical field inversion techniques as described by the manufacturer (FIGE, BioRad). The DNA serves for the preparation of a small size fragment genomic sampling library, i.e. the small-insert library, as well as a large size fragment cluster identification library, i.e. the large-insert library. Both libraries contained DNA fragments generated randomly from genomic DNA and, therefore, they represent the entire genome of *Actinoplanes sp.*

[0097] For the generation of the small-insert library, genomic DNA was randomly sheared by sonication. DNA fragments having a size range between 1.5 and 3 kb were fractionated on a agarose gel and isolated using standard molecular biology techniques (Sambrook et al., Molecular Cloning, 2nd Ed. Cold Spring Harbor Laboratory Press, 1989). The ends of the obtained DNA fragments were repaired using T4 DNA polymerase (Roche) as described by the supplier. This enzyme creates DNA

fragments with blunt ends that can be subsequently cloned into an appropriate vector. The repaired DNA fragments were subcloned into a derivative of pBluescript SK+ vector (Stratagene) which does not allow transcription of cloned DNA fragments. This vector was selected as it contains a convenient polylinker region surrounded by sequences corresponding to universal sequencing primers such as T3, T7, SK, and KS (Stratagene). The unique *Eco*RV restriction site found in the polylinker region was used as it allows insertion of blunt-end DNA fragments. Ligation of the inserts, use of the ligation products to transform *E. coli* DH10B host, selection for recombinant clones, and isolation of plasmids carrying the *Actinoplanes sp.* genomic DNA fragments were performed using well-known methods (Sambrook et al., *supra*). The insert size of 1.5 to 3 kb was confirmed by electrophoresis on agarose gels. Using this procedure a library of small size random genomic DNA fragments is generated that is representative of the entire genome of the studied microorganism. The number of individual clones that can be generated is infinite but only a small number is further analyzed to sample the microorganism's genome.

[0098] To generate the large-insert library, high molecular weight genomic DNA was partially digested with a frequent cutting restriction enzyme, *Sau*3A (G|ATC). This enzyme generates random fragments of DNA ranging from the initial undigested size of the DNA to short fragments of which the length is dependent upon the frequency of the enzyme DNA recognition site in the genome and the extent of the DNA digestion. Conditions generating DNA fragments having an average length of ~40 kb were chosen (Sambrook et al., *supra*). The *Sau*3A restricted DNA was ligated into the *Bam*HI site of the SuperCos-1 cosmid cloning vector (Stratagene) and packaged into phage particles (Gigapack III XL, Stratagene) as specified by the supplier. *E. coli* strain DH10B was used as host and 864 recombinant clones carrying cosmids were selected and propagated to generate the large-insert library. Considering an average size of 8 Mb for an actinomycetes genome and an average size of 35 kb of genomic insert per cosmid in the large insert library, a library of 864 clones represents a 3.78-fold coverage of the microorganism's entire genome. Subsequently, the *Actinoplanes sp.* large-insert library was transferred onto membrane filters (Schleicher & Schnell) as specified by the manufacturer.

[0099] The small insert library was analyzed by sequence determination of the cloned genomic DNA inserts. The universal primers KS or T7, referred to as forward (F) primer, were used to initiate polymerization of labeled DNA. Extension of at least 700

bp from the priming site can be routinely achieved using the TF, BDT v2.0 sequencing kit as specified by the supplier (Applied Biosystems). Sequence analysis of the generated fragments (Genomic Sequence Tags, GSTs) was performed using a 3700 ABI capillary electrophoresis DNA sequencer (Applied Biosystems). The average length of the DNA sequence reads was ~700 bp. Further analysis of the obtained GSTs was performed by sequence homology comparison to various protein sequence databases. The DNA sequences of the obtained GSTs were translated into amino acid sequences and compared to the National Center for Biotechnology Information (NCBI) nonredundant protein database and the proprietary Ecopia natural product biosynthetic gene Decipher™ database using previously described algorithms (Altschul et al., *supra*). Sequence similarity with known proteins of defined function in the database enables one to make predictions on the function of the partial protein that is encoded by the translated GST.

[00100] A total of 882 *Actinoplanes sp.* GSTs were analyzed by sequence comparison. Sequence alignments displaying an E value of at least e-5 were considered as significantly homologous and retained for further evaluation. The E value relates the expected number of chance alignments with an alignment score at least equal to the observed alignment score. An E value of 0.00 indicates a perfect homolog. The E values are calculated as described in Altschul et al. J. Mol. Biol., October 5; 215(3) 403-10. The E value assists in the determination of whether two sequences display sufficient similarity to justify an inference of homology.

[00101] GSTs showing similarity to a gene of interest can be at this point selected and used to identify larger segments of genomic DNA including the gene of interest. Ramoplanins produced by *Actinoplanes sp.* belong to the family of nonribosomal polypeptide antibiotics. Nonribosomal polypeptides are synthesized by nonribosomal peptide synthetase (NRPS) enzymes that perform a series of condensations and modifications of amino acids. Many members of this enzymatic class are found in protein databases rendering possible the identification of an unknown NRPS by sequence similarity. Analysis of the *Actinoplanes sp.* GSTs revealed the presence of three GSTs having similarity to known NRPS proteins in the NCBI nonredundant protein database (Table 1). The obtained E values confirm that these GSTs encode partial NRPS sequences. The three NRPS GSTs were selected for the generation of oligonucleotide probes which were then used to identify gene clusters harboring the specific NRPS genes in the large insert library.

Table 1

	Length (bp)	Proposed function	Homology	Probability	Proposed function of protein match
GST1	632	NRPS	PIR T36248	3.00^{E}-20	CDA peptide synthetase I in *Streptomyces coelicolor*
GST2	592	NRPS	PIR T36248	5.00^{E}-28	CDA peptide synthetase I in *Streptomyces coelicolor*
GST3	502	NRPS	PIR T36180	7.00^{E}-31	CDA peptide synthetase III in *Streptomyces coelicolor*

[00102] Oligonucleotide probes were designed from the nucleotide sequence of the selected GSTs, radioactively labeled, and hybridized to the large-insert library using standard molecular biology techniques (Sambrook et al., *supra*, Schleicher & Schnell). Positive clones were identified, cosmid DNA was extracted (Sambrook et al., *supra*) and entirely sequenced using a shotgun sequencing approach (Fleischmann et al., *Science*, 269:496-512). Identification of the original GSTs, used to generate the oligonucleotide probes, within the DNA sequence of the obtained cosmids confirmed that these cosmids indeed carried the gene cluster of interest.

[00103] Generated sequences were assembled using the Phred-Phrap algorithm (University of Washington, Seattle, USA) recreating the entire DNA sequence of the cosmid insert. Reiterations of hybridizations of the large-insert library with probes derived from the ends of the original cosmid allow indefinite extension of sequence information on both sides of the original cosmid sequence until the complete sought-after gene cluster is obtained. Application of this method on *Actinoplanes sp.* and use of the above-described NRPS GST probes yielded 6 cosmids. Complete sequence of these cosmids and analysis of the proteins encoded by them undoubtedly demonstrated that the gene cluster obtained was indeed responsible for the production of ramoplanin. Subsequent inspection of the ramoplanin biosynthetic cluster sequence, approximately 88.5 kilo base pairs, revealed the presence of three additional GSTs from the small-insert library, bringing the total number of ramoplanin locus GSTs to six.

[00104] Example 2: Genes and Proteins involved in the Biosynthesis of Ramoplanin:

The biological function of the 32 ramoplanin biosynthetic proteins was assessed by computer comparison of each protein with proteins found in the GenBank database of protein sequences (National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD. USA) using the BLASTP algorithm (Altschul et al., 1997, Nucleic Acids Res. Vol. 25, pp.3389-3402). Significant amino acid sequence homologies found for each protein in the ramoplanin locus are shown in Table 2.

Table 2: Proposed functions of the proteins of the ramoplanin biosynthetic pathway based on sequence comparison:

	# aa	proposed function	GenBank accession	probability	% identity	% similarity	proposed function of GenBank match
1	333	unknown; membrane protein	CAB48902	5.00E-22	27	41	possible membrane protein, unknown function, in Streptomyces coelicolor
2	304	ABC transporter	CAB48901	3.00E-55	42	59	probable ABC transporter ATP-binding protein from Streptomyces coelicolor
			AAF81232	7.00E-32	31	47	ABC transporter ATP binding protein found in nonactin biosynthetic locus of Streptomyces griseus
			AAF12291	4.00E-29	34	51	ABC transporter, ATP-binding protein from Deinococcus radiodurans
3	321	unknown; membrane protein	CAB48902	2.00E-15	35	50	possible membrane protein, unknown function, in Streptomyces coelicolor
4	283	oxidoreductase similar to prephenate dehydrogenases	CAA11792	2.00E-69	53	63	similar to prephenate dehydrogenase; chloroeremomycin biosynthesis in Amycolatopsis orientalis
			CAB38592	2.00E-67	50	62	probable oxidoreductase similar to prephenate dehydrogenase; calcium-dependent antibiotic biosynthesis in Streptomyces coelicolor
			AAF67499	3.00E-66	47	64	putative oxidoreductase protein similar to prephenate dehydrogenase; novobiocin biosynthesis in Streptomyces spheroides
5	336	transcriptional regulator similar to StrR	CAA07385	1.00E-74	46	58	StrR DNA-binding protein/regulator of 5'-hydroxystreptomycin biosynthesis in Streptomyces glaucescens; positive transcriptional regulator of strU, strVW genes
			CAB45047	2.00E-74	47	62	probable transcriptional regulator in chloroeremomycin biosynthetic locus of Amycolatopsis orientalis; similar to other regulators of antibiotic biosynthesis
			CAA68515	4.00E-70	47	60	putative regulatory protein StrR in streptomycin biosynthetic locus in Streptomyces griseus
			AAB66654	6.00E-68	44	59	SpcR putative transcriptional regulator of spectinomycin biosynthesis in Streptomyces flavopersicus
			AAF67500	9.00E-58	42	61	NovG putative regulatory protein in novobiocin biosynthetic locus of Streptomyces spheroides
6	444	Amino-transferase	CAB38598	1.00E-123	56	67	possible aminotransferase found in the calcium-dependent antibiotic biosynthetic locus of Streptomyces coelicolor

	# aa	proposed function	GenBank accession	probability	% identity	% similarity	proposed function of GenBank match
			CAA11790	1.00E-101	47	62	protein similar to aminotransferase found in the chloroeremomycin biosynthetic locus of Amycolatopsis orientalis
7	356	oxidoreductase similar to glycolate oxidases	CAB38520	1.00E-115	60	70	putative glycolate oxidase found in calcium-dependent antibiotic biosynthetic locus of Streptomyces coelicolor
			AAA34030	6.00E-77	47	62	spinach glycolate oxidase from Spinacia oleracea
			CAB78838	2.00E-75	45	60	glycolate oxidase-like protein from Arabidopsis thaliana
			CAA11762	4.00E-75	47	61	protein similar to glycolate oxidase in chloroeremomycin biosynthetic locus of Amycolatopsis orientalis
8	640	ABC transporter involved in resistance/transport	CAA11793	0	55	71	protein similar to mdr/ABC transporter found in chloroeremomycin biosynthetic locus of Amycolatopsis orientalis
			AAF67494	1.00E-114	38	57	NovA ABC transporter in novobiocin biosynthetic locus of Streptomyces spheroides
			CAB38879	1.00E-78	34	50	probable ABC transporter found in the calcium-dependent antibiotic biosynthetic locus of Streptomyces coelicolor
9	271	esterase/hydrolase	CAB38877	6.00E-66	48	63	probable hydrolase found in the calcium-dependent antibiotic biosynthetic locus of Streptomyces coelicolor
			CAA11784	9.00E-58	44	56	protein similar to haloperoxidase found in chloroeremomycin biosynthetic locus of Amycolatopsis orientalis
			CAA71338	2.00E-45	41	54	putative thioesterase found in streptothricin biosynthetic locus of Streptomyces sp. strain F20
10	529	unknown	AAB30311	2.00E-29	41	56	unknown protein found in putative chloramphenicol biosynthetic locus of Streptomyces venezuelae
11	90	acyl carrier protein	AAA22001	6.00E-08	33	54	polyketide synthase in Anabaena PCC7120
			CAA98988	8.00E-08	37	57	polyketide synthase found in the phenolpthiocerol biosynthetic locus of Mycobacterium tuberculosis
			AAF62883	7.00E-07	39	55	type I polyketide synthase found in the epothilone biosynthetic locus of Sorangium cellulosum
12	1051	nonribosomal peptide synthetase	CAB15186	0	38	55	nonribosomal peptide synthetase involved in siderophore 2,3-dihydroxybenzoate biosynthesis in Bacillus subtilis
			AAD56240	0	38	55	DhbF peptide synthetase involved in siderophore production in Bacillus subtilis

	# aa	proposed function	GenBank accession	probability	% identity	% similarity	proposed function of GenBank match
			AAC38442	1.00E-179	40	52	actinomycin synthetase II peptide synthetase found in the actinomycin biosynthetic locus of Streptomyces chrysomallus
13	6893	nonribosomal peptide synthetase	AAC80285	0	36	52	SyrE peptide synthetase found in the syringomycin biosynthetic locus of Pseudomonas syringae
			AAC45930	0	31	48	TycC tyrocidine synthetase 3 found in the tyrocidine biosynthetic locus of Brevibacillus brevis
14	8695	nonribosomal peptide synthetase	AAC80285	0	36	51	SyrE peptide synthetase found in the syringomycin biosynthetic locus of Pseudomonas syringae
			AAC45930	0	32	49	TycC tyrocidine synthetase 3 found in the tyrocidine biosynthetic locus of Brevibacillus brevis
15	234	thioesterase	AAC69333	2.00E-30	36	50	PikAV thioesterase II found in the methymycin/pikromycin biosynthetic locus of Streptomyces venezuelae
			AAC01736	6.00E-30	34	49	thioesterase found in the rifamycin biosynthetic locus of Amycolatopsis mediterranei
			CAA57967	2.00E-29	39	48	protein with similarity to thioesterases found in the pyochelin biosynthetic locus of Pseudomonas aeruginosa
			AAA79279	1.00E-28	34	48	thioesterase found in the bialaphos biosynthetic locus of Streptomyces hygroscopicus
16	274	short chain secondary alcohol dehydrogenase/	CAB54559	7.00E-49	39	58	Rhodococcus erythropolis LimC carveol dehydrogenase, a nicotinoprotein belonging to the short chain alcohol dehydrogenase/reductase superfamily
		3-ketoacyl-acyl carrier protein reductase	CAA15546	3.00E-46	39	54	hypothetical protein from Mycobacterium tuberculosis, similar to dehydrogenases
			AAF64503	9.00E-43	39	53	cholesterol oxidase from Nocardioides simplex
			CAA68181	2.00E-38	38	54	UcpA protein, belongs to alcohol dehydrogenase /rybitol dehydrogenase family
			AAC44307	4.00E-36	34	53	FabG 3-ketoacyl-acyl carrier protein reductase from Bacillus subtilis
			CAA77599	1.00E-33	36	49	beta ketoacyl reductase in unknown polyketide biosynthetic locus of Streptomyces cinnamonensis
17	891	threonine-specific adenylate ligase	CAA67248	1.00E-143	49	58	Pristinamycin I synthase 2 nonribosomal peptide synthetase in the pristinamycin biosynthetic locus of Streptomyces pristinaespiralis

	# aa	proposed function	GenBank accession	probability	% identity	% similarity	proposed function of GenBank match
			AAC38442	1.00E-141	49	57	actinomycin synthetase II nonribosomal peptide synthetase in the actinomycin biosynthetic locus of Streptomyces chrysomallus
			CAB38518	1.00E-138	48	58	CDA peptide synthetase I found in the calcium-dependent antibiotic biosynthetic locus of Streptomyces coelicolor
18	187	unknown	none				
19	415	transmembrane protein	CAB42730	2.00E-82	43	57	probable transmembrane protein from Streptomyces coelicolor
			CAB02537	5.00E-59	39	50	probable membrane protein from Mycobacterium tuberculosis
			AAF25828	2.00E-56	35	48	putative transmembrane protein Mycobacterium smegmatis
20	491	halogenase/hydroxylase	CAA11780	1.00E-180	63	76	protein similar to non-heme oxygenase/halogenase found in chloroeremomycin biosynthetic locus of Amycolatopsis orientalis
			CAA76550	1.00E-178	63	75	BhaA protein similar to halogenase, found in the balhimycin biosynthetic locus of Amycolatopsis mediterranei
			AAB49297	1.00E-176	62	74	hypothetical hydroxylase a found in the vancomycin biosynthetic locus of Amycolatopsis orientalis
			AAD24884	6.00E-37	30	46	PltA putative halogenase found in the pyoluteorin biosynthetic locus of Pseudomonas fluorescens
21	217	two-component response regulator	CAB59507	9.00E-58	52	71	Streptomyces coelicolor protein highly similar to various putative two-component response regulators
			CAA22374	8.00E-52	52	66	probable luxR family response regulator from Streptomyces coelicolor
			CAB50960	3.00E-51	49	66	probable two-component system response regulator from Streptomyces coelicolor
			CAB42025	3.00E-48	49	64	probable two-component system regulator from Streptomyces coelicolor
			CAB38597	3.00E-38	44	58	AbsA2, two component response regulator from Streptomyces coelicolor, acts as part of a two component signal transduction system
22	403	two-component sensory protein kinase	CAB42041	1.00E-38	37	48	probable two-component system sensor kinase from Streptomyces coelicolor

	# aa	proposed function	GenBank accession	probability	% identity	% similarity	proposed function of GenBank match
			CAB51250	1.00E-34	32	44	probable two-component system sensor kinase from Streptomyces coelicolor
			CAB89761	1.00E-34	34	42	probable two-component system sensor kinase from Streptomyces coelicolor
			CAB38596	3.00E-27	31	43	AbsA1, two component sensor kinase from Streptomyces coelicolor, acts as part of a two component signal transduction system
23	309	ABC transporter involved in resistance/transport	CAB48901	2.00E-45	41	55	probable ABC transporter ATP-binding protein from Streptomyces coelicolor
			CAB49966	4.00E-28	33	55	ATP-binding transport protein from Pyrococcus abyssi
			AAF12291	9.00E-28	38	56	ABC transporter, ATP-binding protein from Deinococcus radiodurans
24	553	acyl-CoA dehydrogenase	AAD45605	2.00E-18	25	44	isovaleryl-CoA dehydrogenase from Arabidopsis thaliana
			CAB55554	7.00E-18	24	43	isovaleryl-CoA dehydrogenase from Pisum sativum
			CAB46799	4.00E-16	29	44	probable acyl-CoA dehydrogenase from Streptomyces coelicolor
			CAA16488	9.00E-14	29	39	RedW acyl-coa dehydrogenase in the undecylprodigiosin biosynthetic locus of Streptomyces coelicolor
			AAF08800	3.00E-13	23	44	YngJ protein found in the mycosubtilin biosynthetic locus of Bacillus subtilis
25	585	acyl-CoA dehydrogenase	CAB61531	2.00E-27	26	43	FadE fatty acid acyl-CoA dehydrogenase found in Streptomyces lividans
			CAB07077	6.00E-22	24	39	Mycobacterium tuberculosis protein highly similar to acyl-CoA dehydrogenase
			CAA17679	2.00E-21	26	43	probable Acyl-CoA dehydrogenase found in Mycobacterium tuberculosis
26	587	acyl-CoA ligase	AAG02359	1.00E-115	45	56	BlmVI peptide synthetase in bleomycin biosynthetic locus of Streptomyces verticillus
			AAC44128	1.00E-94	38	53	Mx1 peptide synthetase B in saframycin biosynthetic locus of Myxococcus xanthus
			CAA16183	1.00E-85	37	49	polyketide synthase found in the undecylprodigiosin biosynthetic locus of Streptomyces coelicolor
			CAB05426	3.00E-84	35	51	Fad29 probable acyl-CoA synthetase found in Mycobacterium tuberculosis
			CAA17589	2.00E-82	36	51	Fad24 probable acyl-CoA synthetase found in Mycobacterium tuberculosis

	# aa	proposed function	GenBank accession	probability	% identity	% similarity	proposed function of GenBank match
			CAB01395	1.00E-81	35	50	Fad25 probable acyl-CoA synthetase found in Mycobacterium tuberculosis
			AAB52538	2.00E-78	34	50	acyl-CoA synthase from Mycobacterium bovis
			CAB36629	4.00E-78	35	52	putative acyl-CoA synthase from Mycobacterium leprae
27	75	unknown	CAB38589	1.00E-24	70	80	small conserved hypothetical protein found in the calcium-dependent antibiotic biosynthetic locus of Streptomyces coelicolor
			CAB08480	3.00E-22	67	77	MbtH possibly involved in mycobactin synthesis in Mycobacterium tuberculosis
			CAA11799	3.00E-19	74	89	hypothetical protein found in chloroeremomycin biosynthetic locus of Amycolatopsis orientalis
28	94	chorismate mutase-like protein	CAB02002	2.00E-15	50	69	hypothetical protein in Mycobacterium tuberculosis
			CAB82023	2.00E-11	46	59	hypothetical protein in Streptomyces coelicolor
			CAB72783	7.00E-03	36	59	chorismate mutase\prephenate dehydratase from Campylobacter jejuni
			AAC75649	6.00E-02	30	50	chorismate mutase-T and prephenate dehydrogenase protein from E. coli
29	619	membrane protein	CAB16086	2.00E-56	28	43	unknown protein in Bacillus subtilis
			CAA05568	4.00E-34	35	54	YkcB unknown protein in Bacillus subtilis
			CAB76994	0.01	26	35	putative integral membrane protein in Streptomyces coelicolor
			AAC18892	0.049	29	37	transmembrane protein from Streptomyces aureofaciens
30	355	4-hydroxyphenylpyruvate dioxygenase	CAA11761	5.00E-87	50	63	protein similar to hydroxyphenyl pyruvate dioxygenase found in the chloroeremomycin biosynthetic locus of Amycolatopsis orientalis
			CAB38519	1.00E-69	44	54	probable 4-hydroxyphenylpyruvic acid dioxygenase found in the calcium-dependent antibiotic biosynthetic locus of Streptomyces coelicolor
			CAB51008	2.00E-49	36	51	probable 4-hydroxyphenylpyruvic acid dioxygenase found in Streptomyces coelicolor
			AAA50231	3.00E-49	36	50	4-hydroxyphenylpyruvate acid dioxygenase from Streptomyces avermitilis
31	429	transmembrane transporter	CAB45049	4.00E-81	46	64	putative integral membrane ion antiporter found in the chloroeremomycin biosynthetic locus of Amycolatopsis orientalis

	# aa	proposed function	GenBank accession	probability	% identity	% similarity	proposed function of GenBank match
			BAA16991	3.00E-72	39	56	sodium/proton antiporter from Synechocystis sp.
			CAA23036	8.00E-65	37	57	putative sodium/protein exchanging protein from Arabidopsis thaliana
			AAF26906	1.00E-41	30	48	protein similar to sodium/proton and drug/proton antiporters found in the epothilone biosynthetic locus of Sorangium cellulosum
32	189	unknown	CAB72201	1.00E-11	31	41	hypothetical protein in Streptomyces coelicolor
			CAB56690	2.00E-08	31	42	hypothetical protein in Streptomyces coelicolor

[00105] The correlation between the order of repeated units in most peptide synthetases and the order in which the respective amino acids appear in the peptide product makes it possible to correlate peptides of known structure with putative genes encoding their synthesis, as demonstrated by the identification of the mycobactin biosynthetic gene cluster from the genome of Mycobacterium tuberculosis (Quadri et al., 1998, Chem. Biol. Vol. 5, pp. 631-645). This principle has been used here to assign a biosynthetic role for each repeating unit of the ramoplanin peptide synthetases described in this invention, as diagrammed in Figure 2A, B and C. The approximate boundaries, at the amino acid level, of the domains of the repeating units (modules) of each ORF are tabulated in Table 3, wherein C represents a condensation domain, A represents an adenylation domain, T represents a thiolation domain and Te represents a thioesterase domain.

Table 3: Approximate boundaries of domains of each moldule at the amino acid level

Orf 12		
Module 1:	C	1-470
	A	471-959
	T	961-1030
Orf 13		
Module 1:	C	1-517
	A	518-990
	T	991-1059
Module 2:	C	1106-1560
	A	1561-2052
	T	2054-2122
Module 3:	C	2159-2618
	A	2619-3122
	T	3123-3191
Module 4:	C	3237-3697
	A	3698-4160
	T	4161-4228
Module 5:	C	4241-4718
	A	4719-5192
	T	5193-5260
Module 6:	C	5307-5754
	T	5755-5824
Module 7:	C	5838-6317
	A	6318-6804
	T	6805-6873
Orf 14		
Module 1:	C	1-486
	A	487-993
	T	994-1062
Module 2:	C	1109-1567
	A	1568-2041
	T	2042-2110
Module 3:	C	2122-2602
	A	2603-3095
	T	3097-3165
Module 4 :	C	3212-3671
	A	3672-4135
	T	4136-4202
Module 5 :	C	4217-4698
	A	4699-5199
	T	5200-5268
Module 6 :	C	5317-5776
	A	5777-6280
	T	6281-6350
Module 7 :	C	6363-6839
	A	6840-7343
	T	7344-7411
Module 8 :	C	7458-7925
	A	7926-8380
	T	8381-8449
	Te	8450-8695

[00106] A. Formation of the lipodepsipeptide core structure:

Nine proteins, encoded by ORFs 9, 11, 12, 13, 14, 15, 17, 26 and 27 (SEQ ID NOS: 10, 12, 13, 14, 15, 16, 18, 27 and 28), are likely to be involved in the formation of the lipodepsipeptide core structure of ramoplanin. ORFs 11, 12, 13, 14 and 17 (SEQ ID NOS: 12, 13, 14, 15 and 18) show significant similarity to peptide synthetases or

peptide synthetase domains. Analysis of the adenylation domains found in these ORFs allows the amino acid that is incorporated by each unit to be identified (see Figure 3 A and B). The following amino acid specificities are consistent with these comparisons: ORF 12: asparagine (Asn); ORF 13, module 1: 4-hydroxyphenylglycine (HPG); ORF 13, module 2: ornithine (Orn); ORF 13, module 3: threonine (Thr); ORF 13, module 4: HPG; ORF 13, module 5: HPG; ORF 13, module 6 contains no adenylation domain; ORF 13, module 7: phenylalanine (Phe); ORF 14, module 1: Orn; ORF 14, module 2: HPG; ORF 14, module 3: Thr; ORF 14, module 4: HPG; ORF 14, module 5: glycine (Gly); ORF 14, module 6: leucine (Leu); ORF 14, module 7: unspecified; ORF 14, module 8: HPG; ORF 17, threonine (Thr). The numbers and predicted amino acid substrate specificities of the peptide synthetase repeating units are in precise agreement with the structure of the ramoplanin peptide core, providing conclusive evidence that the genetic locus described here is responsible for the biosynthesis of ramoplanin.

[00107] The amino acid specificity of adenylation domains may be altered by mutagenesis (Stachelhaus et al., 1999, Chem. Biol. Vol. 6, pp. 493-505; Challis et al., Chem. Biol., 2000, Vol. 7, pp. 211-224) or by swapping domains between peptide synthetases (Stachelhaus et al., 1995, Science Vol. 269, pp. 482-485; Schneider et al., 1998, Mol. Gen. Genet. Vol. 257, pp. 308-318; de Ferra et al., 1998, J. Biol. Chem. Vol. 272, pp. 25304-25309) and thereby generate derivatives of a natural peptide product.

[00108] A model for the biosynthesis of the ramoplanin peptide core structure can be built by comparing the specificity and order of the repeating units in the ramoplanin peptide synthetases with the order of the amino acid substituents in ramoplanin (diagrammed in Figure 2A and C). ORF 12 (SEQ ID NO: 13) contains the only adenylation domain specifying Asn and therefore may catalyze the incorporation of the first two (Asn) amino acid residues into the peptide chain. Subsequent amino acids are incorporated in the precise order in which the respective units occur in the adjacent ORFs 13 and 14 (SEQ ID NOS: 14 and 15) . The only exception to the colinearity of peptide synthetase units and the order of incorporation of amino acids into ramoplanin occurs at module 6 of ORF 13 (SEQ ID NO: 14). This module contains condensation and thiolation domains, but is lacking an adenylation domain. The structure of ramoplanin indicates that a Thr must be incorporated into the peptide chain at this position. ORF 17 (SEQ ID NO: 18) encodes an unusual peptide synthetase unit having an adenylation domain that specifies Thr, but lacks a conventional condensation

domain. According to the model diagrammed in Figure 2A, the ORF 17 (SEQ ID NO: 18) protein interacts with module 6 of ORF 13 (SEQ ID NO: 14) and substitutes for the missing adenylation domain of this module, thus catalyzing the incorporation of Thr into the growing ramoplanin peptide precursor at the appropriate position. Such a *trans* interaction between peptide synthetase units has a precedent in the biosynthesis of the lipodepsipeptide antibiotic syringomycin. In the syringomycin system, the adenylation domain of the SyrB1 protein, which lacks a condensation domain, is proposed to interact with and complement the activity of a SyrE1 peptide synthetase unit that contains a condensation domain but is lacking an adenylation domain (Guenzi et al., 1998, J. Biol. Chem. Vol. 273, pp. 32857-32863).

[00109] The peptide synthetase encoded by ORF 12 (SEQ ID NO: 13) is unusual for a starter unit in having a condensation domain at the N-terminus of the protein. Most peptide synthetase starter units described to date contain adenylation domains at their N-terminus that are responsible for activating the first amino acid (the "starter" amino acid) that is incorporated into the peptide product. In contrast, the ramoplanin starter unit encoded in ORF 12 (SEQ ID NO: 13) has a condensation domain at the N-terminus of the protein, indicating that the initiation of peptide synthesis may occur in an unusual fashion. The N-terminus of the ramoplanin peptide is modified by one of three possible fatty acid groups, suggesting that the construction of the ramoplanin peptide may start with a fatty acid rather than an amino acid. A proposed mechanism of chain initiation using a fatty acid starter group is diagrammed in Figure 2B. According to this model, the condensation domain at the N-terminus of ORF 12 (SEQ ID NO: 13) catalyzes the linkage of amino acid 1 (Asn) bound to module 1 to a fatty acid bound to the acyl carrier protein encoded by ORF 11 (SEQ ID NO: 12) via amide bond formation, providing an "acyl-N-capped" amino acid intermediate for further chain extension.

[00110] ORFs 11 and 26 (SEQ ID NOS: 12 and 27) are proposed to cooperate in the activation and transfer of fatty acid precursors to the ORF 12 (SEQ ID NO: 13) peptide synthetase. ORF 26 (SEQ ID NO: 27) shows similarity to acyl-CoA ligases, proteins of the adenylate-forming superfamily of enzymes that catalyze the activation of fatty acids via an activated adenylate intermediate. ORF 11 (SEQ ID NO: 12) shows similarity to acyl carrier proteins and peptide synthetase thiolation domains that accept activated adenylate intermediates. As diagrammed in Figure 2B, the activity of these two ORFs may generate activated fatty acid thioesters that serve as the initiating groups for the synthesis of the ramoplanin lipopeptide core structure. ORF 26 (SEQ ID NO: 27) may

be replaced or mutated, alone or in combination with the condensation domain of ORF 12 (SEQ ID NO: 13), in order to generate derivatives of ramoplanin having alternative fatty acids.

[00111] The final unit in most peptide synthetases contains a special C-terminal thioesterase domain, postulated to be involved in product release. Release of the complete peptide product from the peptide synthetase requires a thioesterase function that is generally found at the C-terminus of the peptide synthetase. ORF 14 (SEQ ID NO: 15) contains a C-terminal thioesterase domain, and may be involved in peptide release and cyclization by catalyzing the formation of the ester bond between the carboxylate goup of the C-terminal HPG and the hydroxyl group of HAsn, resulting in a free cyclic depsipeptide structure. ORF 15 (SEQ ID NO: 16) may also play a role in peptide release and/or cyclization. ORF 15 (SEQ ID NO: 16) shows strong similarity to thioesterases that are frequently found associated with peptide synthetases and are postulated to be involved in the release of peptide products or intermediates and may also be involved in the release and/or cyclization of the ramoplanin peptide. ORF 9 (SEQ ID NO: 10) shows similarity to esterases of the alpha/beta hydrolase fold family and may also be involved in peptide release.

[00112] ORF 27 (SEQ ID NO: 28) shows strong similarity to several small conserved proteins encoded by genes that are frequently found to be associated with peptide synthetase genes and are therefore likely to play a role in peptide biosynthesis.

[00113] B. Epimerization of L-amino acids into corresponding D-amino acids:

An unexpected feature of the ramoplanin peptide synthetases is the absence of epimerization domains in the repeating units. Epimerization domains catalyze the conversion of L-amino acids into the corresponding D-amino acids. Ramoplanin contains seven D-amino acid units. Most bacterial peptide synthetases that incorporate D-amino acids do so by first recognizing and incorporating the corresponding L-amino acid and subsequently altering the configuration to the D- form through the activity of the epimerization domain. The lack of epimerization domains in the ramoplanin peptide synthetases despite the presence of D-amino acids in the final natural product may be due to specific recognition of D-amino acids by the adenylation domains found in modules 1, 2, 3 and 5 of ORF 13 (SEQ ID NO: 14) and modules 1, 3 and 7 of ORF 14 (SEQ ID NO: 15). The direct recognition and incorporation of D-amino acids by peptide synthetases has been postulated for the eukaryotic cyclosporin and HC toxin peptide

synthetases (Weber et al., 1994, Curr. Genet Vol. 26, pp. 120-125; Scott-Craig et al., 1992, J. Biol. Chem. Vol. 267, pp. 26044-26049).

[00114] Alternatively, epimerization may be catalyzed by cellular amino acid epimerases/epimerases of primary or secondary metabolism, as has been proposed for the incorporation of D-valine in the gramicidin and tyrocidine systems (Pfeifer et al., 1995, Biochem. Vol. 34, pp. 7450-7459; Stein et al., 1995, Biochem. Vol. 34, pp. 4633-4642).

[00115] Yet another explanation is that specialized domains within the NRPSs may have evolved the ability to carry out dual functions. One domain that stands out as a candidate for having such dual functions is the condensation domain. Normally within a typical NRPS module that introduces a D-amino acid into the peptide product, epimerization (E) domains follow the thiolation (T) domain. In terms of linear domain organization on NRPS enzymes condensation (C) domains and epimerization (E) domains can be thought of occupying equivalent positions. That is, in an NRPS with multiple modules that is devoid of E domains, a C domain from any given module is found directly adjacent to the thiolation (T) domain of the upstream module. In addition, C domains and E domains also share a considerable amount of sequence similarity. Several highly conserved core motifs are shared between C and E domains. One particularly important motif that is common to both C and E domains is the histidine motif HHXXXDG which has been shown by mutagenesis to form part of the active site (Stachelhaus *et al.*; *Journal of Biological Chemistry* 1998;273:22773-22781). Thus, the C domains of modules 2, 3, 4 and 6 of OFR 13 (SEQ ID NO:14) and modules 2, 4 and 8 of ORF 14 (SEQ ID NO: 15) may be capable of amino acid epimerization as well as amide bond formation and be responsible for the 7-D-amino acid residues found in ramoplanin.

[00116] C. Formation of fatty-acid side chains:

The ramoplanin depsipeptide core structure may carry one of three different medium-chain fatty acids attached to the N-terminus of Asn in position 1, resulting in the three different ramoplanin components A1-A3. Little is known about the biosynthetic origin of the three unsaturated fatty acid precursors, octa-2,4-dienoic acid (a C8 fatty acid) and its analogs 7-methylocta-2,4-dienoic acid (C9) and 8-methylnona-2,4-dienoic acid (C10). These medium-chain fatty acids may be derived from longer chain fatty acids by beta-oxidative degradation. It has been shown that the yields of

component A2, carrying the octa-2,4-dienoic acid moiety, can be increased by adding the amino acid leucine to the fermentation medium of the producing organism, indicating that branched-chain amino acids may also serve as biosynthetic precursors to the fatty acid side chains of ramoplanin (European patent EP259780). Three proteins encoded by the ramoplanin locus, namely ORFs 16, 24, 25 (SEQ ID NOS: 17, 25 and 26), show similarity to enzymes associated with fatty acid metabolism and therefore may be involved in the generation of the fatty acid side chains for attachment to the depsipeptide core structure of ramoplanin. ORFs 24 and 25 (SEQ ID NOS: 25 and 26) are highly similar to each other and to flavin-dependent acyl-CoA dehydrogenases, enzymes involved in the degradation of fatty acids and in the degradation of leucine to fatty acid intermediates. These ORFs may channel branched-chain amino acid and fatty acid intermediates into the ramoplanin biosynthetic pathway. In addition, the dehydrogenase activity of ORFs 24 and 25 (SEQ ID NOS: 25 and 26) may be responsible for generating the two double bonds found in the unsaturated fatty acid groups of ramoplanin. ORF 16 (SEQ ID NO: 17) may also be involved in generating the fatty acid group of ramoplanin as it shows strong similarity to 3-oxoacyl-acyl carrier protein reductases, NAD-dependent enzymes of primary metabolism that are also involved in fatty acid degradation.

[00117] D. Amino-acid 4-hydroxyphenylglycine (HPG) synthesis:

Five proteins encoded by the ramoplanin locus, namely ORF 4, ORF 6, ORF 7, ORF 28 and ORF 30 (SEQ ID NOS: 5, 7, 8, 29 and 31), are likely to be involved in synthesizing the unusual amino acid 4-hydroxyphenylglycine (HPG) which serves as a substrate for incorporation into the lipodepsipeptide core structure of ramoplanin. The natural occurrence of HPG in secondary metabolites is relatively infrequent, the best-known examples being nocardicin A; vancomycin, aridicin, chloroeremomycin, teicoplanin and related glycopeptide antibiotics; the calcium-dependent antibiotic (CDA) of Streptomyces coelicolor; and ramoplanin. Biochemical studies have indicated that the HPG residues of the antibiotics vancomycin, aridicin, and nocardicin are derived from the common amino acid tyrosine and a pathway for the synthesis of HPG from tyrosine has been proposed (Nicas et al., in Biotechnology of Antibiotics, Marcel Dekker, Inc., 1997, pp. 363-392 and references therein; Chung et al., 1986, J. Antibiotics Vol. 1986, pp. 642-651; Hosoda et al., 1977, Agric. Biol. Chem. Vol. 41, pp. 1007-1012; Hammond et al., 1982, J. Chem. Soc. (Chem. Comm.), Vol. 1982, pp. 344-

346). However, analysis of the ORFs encoded by the ramoplanin biosynthetic locus provides evidence for an alternative pathway, as illustrated in Figure 4. The combined activities of ORF 4, ORF 6, ORF 7, ORF 28 and ORF 30 (SEQ ID NOS: 5, 7, 29, and 31) would allow conversion of intermediates of tyrosine metabolism into the unusual amino acid HPG. Proteins showing similarity to ORFs 4, 6, 7 and 30 (SEQ ID NOS: 5, 7, 8 and 31) can be found in the biosynthetic loci encoding CDA and chloroeremomycin, two natural products that also contain HPG substituents, although the roles of these proteins in the biosynthesis of the respective natural products were not proposed (GenBank accession numbers AL035640, AL035707, and AL035654; van Wageningen et al. 1997, Chem. Biol. Vol. 5, pp. 155-162).

[00118] E. Resistance and/or localization proteins:

Eight proteins encoded by the ramoplanin locus (ORF 1, ORF 2, ORF 3, ORF 8, ORF 19, ORF 23, ORF 29 and ORF 31) are likely to be membrane-associated proteins that are involved in resistance and/or the subcellular localization of the ramoplanin biosynthetic machinery. ORFs 2, 8, and 23 (SEQ ID NOS: 3, 9 and 24) show similarity to the superfamily of ATP binding cassette transport proteins involved in target-specific secretion and are likely to be involved in the transport of ramoplanin or biosynthetic precursors across the cytoplasmic membrane, providing a possible mechanism for resistance to the toxic effects of the antibiotic or increased production of ramoplanin. ORF 31 (SEQ ID NO: 32) shows similarity to various sodium/proton and drug/proton antiporters and may also provide a means to transport ramoplanin across the cytoplasmic membrane. ORFs 1, 3, 19 and 29 (SEQ ID NOS: 2, 4 and 20) show similarity to various transmembrane proteins of unknown function and may be involved in localizing the ramoplanin biosynthetic machinery to the cytoplasmic membrane in order to provide access to lipid and fatty acid precursors.

[00119] F. Proteins involved in regulation of ramoplanin biosynthesis:

Three proteins encoded by the ramoplanin locus, namely ORF 5, ORF 21, ORF 22 (SEQ ID NOS: 6, 22 and 23), are likely to be involved in the regulation of ramoplanin biosynthesis. ORF 5 (SEQ ID NO: 6) shows similarity to a number of transcriptional regulators of antibiotic biosynthesis. This protein is likely to regulate the transcription of one or more genes in the ramoplanin genetic locus. ORFs 21 and 22 (SEQ ID NOS: 22 and 23) show homology to 2-component signal transduction systems, such as the Abs

A1/ A2 system involved in the global regulation of antibiotic synthesis of *Streptomyces coelicolor*. These ORFs may act coordinately to regulate the expression of ramoplanin biosynthetic genes and the production of ramoplanin in response to environmental or cellular signals.

[00120] G. Chlorination of terminal HPG residue:

ORF 20 (SEQ ID NO: 21) shows similarity to halogenases involved in the chlorination of secondary metabolites, including the PrnC halogenase of *Pseudomonas fluorescens* responsible for the chlorination of an aromatic precursor of pyrrolnitrin biosynthesis and a halogenase proposed to be responsible for the chlorination of a tyrosine residue in chloroeremomycin. This protein most likely catalyzes the chlorination of the terminal HPG residue incorporated into the ramoplanin peptide core, generating the 3-chloro-HPG form.

[00121] H. Beta-hydroxyasparagine residue formation:

As disclosed in USSN 60/283,296, ORF 10 (SEQ ID NO: 11) is a member of a new family of metal cofactor hydroxylase enzymes. This discovery is very surprising because one would have expected that cytochrome P450 enzymes would be implicated in the beta-hydroxylation reaction requied to generate beta-hydroxyasparagine.

[00122] The possibility that a novel mechanism for beta-hydroxylation of amino acid residues may be operative in the biosynthesis of ramoplanin was first suggested by the fact that none of the ORFs encoded by the ramoplanin biosynthetic locus displayed significant amino acid sequence homology to the known cytochrome P450 monooxygenases by BLASTP analysis. ORF 10, ORF 18 and ORF 32 (SEQ ID NOS: 11, 19 and 33) could not initially be assigned a putative role in the biosynthesis of ramoplanin and were considered as candidate asparagine beta-hydroxylases. ORF 10 (SEQ ID NO: 11) shows homology to a protein of unknown function in the bleomycin biosynthetic locus of *Streptomyces verticillus* and to a partial protein of unknown function found in putative chloramphenicol biosynthetic locus of *Streptomyces venezuelae*. Significantly, bleomycin and chloramphenicol also contain a beta-hydroxylated amino acid residue. ORF 18 (SEQ ID NO: 19) shows no similarity to proteins in the GenBank database, while ORF 32 (SEQ ID NO: 33) shows similarity to hypothetical bacterial proteins of unknown function in *Streptomyces coelicolor*. Since enzymes that catalyze hydroxylation reactions commonly use metal cofactors, ORFs

10, 18 and 32 (SEQ ID NOS: 11, 19 and 33) were further analyzed for the presence of amino acid motifs that are associated with the binding of metal cofactors.

[00123] Figure 5 illustrates clustal alignments showing sequence homology between ORF 11 (SEQ ID NO: 12) and various metal ligand motifs. In each of the clustal alignments: (i) a line above the alignment is used to mark strongly conserved positions; (ii) an asterisk "*" indicates positions which have a single, fully conserved residues; (iii) a colon ":" indicates that one of the following strong groups is fully conserved: STA; NEQK; NHQK; NDEQ; QHRK; MILV; MILF; HY; and FYW; and (iv) a period "." indicates that one of the following weaker groups is fully conserved: CSA; ATV; SAG; STNK; STPA; SGND; SNDEQK; NDEQHK; NEQHRK; FVLIM: and HFY.

[00124] ORF 10 (SEQ ID NO: 11) contains two amino acid sequence motifs that are frequently found in enzymes that use metal cofactors. The N-terminal region of ORF 10 (SEQ ID NO: 11) contains a cluster of histidine residues (the His-motif) that shows significant local sequence homology to a conserved histidine motif found in several zinc-binding beta-lactamases. Figure 5A shows the local amino acid sequence homology between ORF 10 (SEQ ID NO: 11) and a key motif involved in coordinating two zinc molecules in the beta-lactamase superfamily. The alignment depicts amino acids 263 to 318 of ORF 10 (SEQ ID NO: 11), amino acids 42 to 99 of a member of the beta-lactamase superfamily, the L1 metallo-beta-lactamase (1SML) from *Stenotrophomonas maltophilia* for which the crystal structure has been determined (Ullah *et al.,* 1998), and amino acids 12 to 67 of the consensus sequence for pfam00753, i.e. the beta-lactamase superfamily motif (Bateman *et al.*, 2000). Highlighted in black are residues demonstrated in the L1 metallo-beta-lactamase to co-ordinate zinc and their counterparts in the other two sequences. X-ray crystal structure analysis demonstrates that the histidine residues in this conserved motif are responsible for binding the zinc metal cofactor (Ullah *et al.,* 1998). The precise alignment and conserved spacing of the amino acid residues in the His-motif of ORF 10 (SEQ ID NO: 11) as compared to the zinc-binding beta-lactamases indicates that ORF 10 (SEQ ID NO: 11) is likely to bind a metal cofactor.

[00125] Figure 5B shows the local amino acid sequence homology between ORF 10 (SEQ ID NO: 11) and a key motif involved in coordinating an iron molecule in cytochrome P450 monooxygenases. The alignment depicts amino acids 405 to 452 of ORF 10 (SEQ ID NO: 11) and amino acids 370 to 421 of the consensus sequence for pfam00067, i.e. the cytochrome P450 motif (Bateman *et al.*, 2000). The region of ORF

10 (SEQ ID NO: 11) in highlight is in relatively good agreement with the Prosite motif PS00086 (Hofmann *et al.*, 1999) required for binding iron: [FW]-[SGNH]-x-[GD]-x-[RHPT]-x-C-[LIVMFAP]-[GAD], where x is any amino acid and amino acids in brackets indicate the variability in a given position. Notably, the least variable positions of this motif are present in ORF10 (SEQ ID NO: 11), i.e. residues Phe-423, Gly-425, Cys-428, and Gly-430). The C-terminal region of ORF 10 (SEQ ID NO: 11) contains a cluster of amino acid residues that shows significant local sequence homology to a motif frequently found in cytochrome P450 monooxygenases (the Cys-motif). This motif includes a cysteine residue that is highly conserved in the cytochrome P450 monooxygenases and that has been shown by X-ray crystal structure analysis to be involved in binding the iron metal cofactor required for catalysis. The Cys-motif of ORF 10 (SEQ ID NO: 11) is likely to contribute to the binding of a metal cofactor. The presence of two amino acid sequence motifs that are found in well-characterized metal-binding enzymes indicates that ORF 10 (SEQ ID NO: 11) is likely to be a metal-binding enzyme. Thus, the ORF 10 (SEQ ID NO: 11) is likely to be responsible for the formation of beta-hydroxyasparagine during the synthesis of ramoplanin.

[00126] Example 3: Expression analysis

A – Acyl starter unit chain initiation

To investigate the involvement of an acyl starter unit chain in chain initiation of the ramoplanin NRPS system, ORF 11, ORF 12, and ORF 26 (SEQ ID NOS: 12 to 14) were individually PCR-amplified using oligonucleotide primer pairs that introduced convenient restriction enzyme sites at either end of each ORF as well as ten consecutive histidine tags at the N-terminus. These recombinant N-terminal HIS10-tagged ORFs were subcloned into an *E. coli* expression vector and the resulting plasmids were introduced into *E. coli* which were then grown under conditions that lead to high level expression of the recombinant ORFs. Cells were pelleted and disrupted, and the recombinant ORF 11, ORF 12, and ORF 26 (SEQ ID NOS: 12, 13 and 27) proteins were purified by nickel affinity chromatography. The ORF 11 and ORF 26 (SEQ ID NOS: 12 and 27) proteins are readily obtained as soluble protein preparations whereas the solubility of ORF 26 (SEQ ID NO: 27) is more reduced presumably due to its large size.

[00127] Based on sequence homology, ORF 11 (SEQ ID NO: 12) is predicted to be an acyl or amino acyl carrier protein. Purified recombinant ORF 11 (SEQ ID NO: 12)

protein can be primed to its holo form *in vitro* using purified Sfp from *Bacillus subtilis* and coenzyme A, as indicated by an increase in mass by MALDI-MS that corresponds to the addition of the 4'-phosphopantetheine moiety of coenzyme A. The fact that recombinant ORF 11 is amenable to this posttranslational modification that converts it from an inactive apo into the active holo form confirms that it is indeed an acyl or amino acyl carrier protein.

[00128] The availability of solube recombinant ORF 26 together with solube, holo ORF 11 (described above) provides a means to confirm ORF 26's role in the transfer of the short chain fatty acids onto holo ORF 11. Such an experiment using as substrate the ^{14}C-radiolabeled long chain fatty acid palmetic acid was inconclusive. These findings are consistent with the hypothesis that ORF 26 is specific for shorter chain fatty acids such as the three 8- to 10-carbon unsaturated fatty acids found in ramoplanins rather than long chain saturated fatty acids such as 16-carbon palmitic acid. Substrate specificity is further examined by synthesis of the fatty acyl groups that are naturally found linked to the amino terminus of the ramoplanin peptide.

[00129] B – beta-hydroxyasparagine

To confirm characterization of ORF 10 (SEQ. ID NO: 11) as a beta-hydroxylase and to confirm the role of ORF 10 (SEQ. ID NO: 11) in hydroxylation of asparagine at the beta position, a recombinant N-terminal His10-tagged ORF 10 *E. coli* expression system was designed as described above for ORFs 11, 12 and 26 (SEQ ID NOS: 12, 13 and 27). Purified recombinant ORF 10 (SEQ ID NO: 11) protein was obtained in a soluble form by nickel affinity chromatography. The fact that the purified recombinant protein does not display the characteristic absorption spectrum of heme-containing enzyme indicates that ORF 10 (SEQ ID NO: 11) is not a P450 enzyme. The ORF 10 (SEQ ID NO: 11) metal-binding motifs mentioned above therefore co-ordinate a non-heme iron or a metal other than iron.

[00130] As an alternative source of native ORF 10 (SEQ ID NO: 11), a *Streptomyces* expression system was employed. ORF 10 (SEQ ID NO: 11) was amplified by high fidelity PCR using two specific oligonucleotides, namely primer sequences (5' to 3') N-oligo: CACACAGAATTCACCAGCGCCACTCGCGCTT, and C-oligo: CACACATCGATGGGCAACGCCGATCAGCCG. This primer pair introduces convenient restriction enzyme sites at either end of the ORF 10 gene but does not introduce any exogenous amino acids. The amplified genes were then subcloned using

*Cla*I and *Eco*RI restriction enzymes into a *Streptomyces*/E.coli expression shuttle vector, pECO1202. Following confirmation of the cloned sequences, *Streptomyces lividans* TK24 was transformed with this construct. Five independent transformants were selected for further analysis. Cultures were grown for 48 hours in a gyrating 30°C incubator using 25 ml erlenmeyer flasks containing 5 ml of Tryptic Soy Broth (TSB, Difco). Total RNA was extracted from the cell pellets using the RNeasy kit (Qiagen). The integrity and concentration of the RNA was monitored by agarose gel electrophoresis. Subsequently, reverse transcription was performed using 1 ug total RNA primed with an antisense primer sequence located in the vector just downstream of the stop codon. Following reverse transcription of each sample and appropriate controls, 20 cycles of PCR were performed using the original ORF-specific oligonucleotides, N-oligo and C-oligo. According to the RT-PCR analysis, the five recombinant *S. lividans* clones express relatively high levels of ORF 10-specific mRNA and the size of the RT-PCR product is as expected. Figure 6 shows the RT-PCR analysis of recombinant *S. lividans* clones expressing ramoplanin ORF 10, wherein is lane 1 is 1 kb DNA ladder; lane 2 is untransformed *S. lividans*; lane 3 is *S. lividans* transformed with empty expression vector; lanes 4-8 are five different *S. lividans* recombinant clones expressing ramoplanin orf 10; lane 9 is an *S. lividans* recombinant clone expressing an unrelated gene; lane 10 is negative control performed without RNA; lane 11 is negative control performed without RT; lane 12 is positive control for PCR using plasmid DNA.

[00131] To confirm that these recombinant strains actually produce the expected ORF 10 protein lysates were analyzed by SDS-PAGE. Briefly, cell pellets from the above cultures were resuspended in cold extraction buffer (0.1M Tris-HCl, pH 7.6, 10mM $MgCl_2$, 1mM PMSF) and sonicated four times for 20 sec on ice with 1 min intervals. Soluble proteins were recovered by centrifugation for 10 min at 20, 000 X *g* and the total protein concentration was determined using the Bradford reagent (Biorad). Equal amounts of total soluble protein were subjected to 10% SDS-PAGE analysis. Proteins were visualized by staining with coomassie brilliant blue.

[00132] As shown in Figure 7, the four recombinant strains tested contain a significant amount of protein with an apparent mobility of approximately 60 kilodaltons, consistent with the predicted molecular mass of 58916.80 kilodaltons for the ORF 10 protein. Figure 7 is the SDS-PAGE analysis of recombinant *S. lividans* clones expressing ramoplanin ORF 10 (SEQ ID NO.:11). The soluble fraction of protein

lysates was subjected to 10% SDS-PAGE and stained with coomassie blue. Lane 1 is molecular weight standards with sizes in kilodaltons indicated to the left; lane 2 is untransformed *S. lividans*; lane 3 is *S. lividans* transformed with empty expression vector; lanes 4 to 7 are four different *S. lividans* recombinant clones expressing ramoplanin ORF 10 (SEQ. ID NO.:11). The approximately 60kDa ORF 10 gene product is clearly visible in lanes 4 to 7, as indicated by the arrowhead to the right.

[00133] It is to be understood that the embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents and patent applications and sequences from GenBank and other databases referred to herein are incorporated by reference in their entirety for all purposes.